REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
4 August 1998

3. REPORT TYPE AND DATES COVERED

4. TITLE AND SUBTITLE
THE EFFECTS OF HIGHLIGHTING, VALIDITY, AND FEATURE TYPE ON AIR-TO-GROUND TARGET ACQUISITION PERFORMANCE

5. FUNDING NUMBERS

6. AUTHOR(S)
Rena Adria Conejo

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Illinois at Urbana-Champaign

8. PERFORMING ORGANIZATION REPORT NUMBER
98-034

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
THE DEPARTMENT OF THE AIR FORCE
AFIT/CIA, BLDG 125
2950 P STREET
WPAFB OH 45433

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION AVAILABILITY STATEMENT
Unlimited distribution
In Accordance With AFI 35-205/AFIT Sup 1

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

14. SUBJECT TERMS

15. NUMBER OF PAGES
60

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT

THE EFFECTS OF HIGHLIGHTING, VALIDITY, AND FEATURE TYPE ON AIR-TO-GROUND TARGET ACQUISITION PERFORMANCE

2LT Rena Adria Conejo, USAF
M.S., Department of Psychology
University of Illinois at Urbana-Champaign, 1997
Dr. Christopher D. Wickens, Advisor

Air-to-ground target acquisition is an important part of any flight task, whether during navigation, landing at an airport, or during military bombing missions. This task usually involves visual searches to cross-check between a map and the forward field of view (FFOV). Previous research has implied that cultural (man-made) targets and lead-in features may yield better performance than natural targets and lead-in features. Also, since the target acquisition task often involves visual searches in complex visual fields (for both the map and the FFOV), previous research suggests that highlighting a target on the map can automate the search process by increasing the target's salience. However, as with any automation, the possibility of failure exists and is often associated with drastic consequences. This study examined the differences in performance between natural and cultural feature types as targets and lead-in features under highlighted and nonhighlighted conditions. Also, performance under highlighted and nonhighlighted conditions were compared to determine if highlighting did facilitate the target acquisition task. When targets were highlighted, the highlighting was either valid, invalid with the wrong target highlighted, or invalid with the wrong target highlighted and the correct target absent from the FFOV. The target highlighting occurred under 60% validity conditions; therefore, in some cases, a nearby lead-in feature was highlighted (always validly) to minimize or eliminate any costs of highlighting invalidity. Analysis showed that performance according to feature type was best under a target by lead-in interaction where the target feature type was opposite that of the lead-in feature type. Valid highlighting did not provide any significant benefit over nonhighlighted conditions; yet invalid target highlighting produced performance costs that were not sufficiently improved by any feature-type interaction or the lead-in highlighted condition. Further analyses suggest that performance under target absent conditions may result from different cognitive processes than those involved in conditions where the target is in the FFOV. Hence, target highlighting is not recommended unless accuracy of highlighting can be guaranteed. Also, further study must determine if the beneficial target by lead-in feature interaction reported here is generic to all tasks and environments.

References

Antunano, M.J., Mohler, S.R., and Gosbee, J.W. (1989). Geographic disorientation: Approaching and landing at the wrong airport. *Aviation, Space, & Environmental Medicine,* **55**, 996-1104.

Aretz, A.J. (1991). The design of electronic map displays. *Human Factors,* **33**(1), 85-101.

Battiste, V., and Delzell, S. (1991). Visual cues to geographical orientation during low-level flight. *Proceedings of the 6th International Symposium on Aviation Psychology.* Columbus, OH: Department of Aviation, Ohio State University.

Biederman, I. and Gerhardstein, P.G. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance,* **19**(6), 1162-1182.

Brown, T.J. (1991). Visual display highlighting and information extraction. *Proceedings of the Human Factors Society 35th Annual Meeting.* Santa Monica, CA: Human Factors Society.

Carmody-Bubb, M.A., and Dunn, R.S. (1996). *Situational awareness and mission rehearsal training systems: An implicit approach.* Patuxent River, ML: Naval Air Warfare Center.

Carter, R.C., and Cahill, M.C. (1979). Regression models of search time for color coded information displays. *Human Factors,* **21**(3), 293-302.

Christner, C.A. and Ray, H.W. (1961). An evaluation of the effect of selected combinations of target and background coding on map-reading performance -- experiment V. *Human Factors,* **3**, 131-146.

Converse, S.A., Kozai, S., and Batten, D. (1992). Color coding to facilitate performance of focused attention tasks with object displays. *Proceedings of the Human Factors Society 36th Annual Meeting.* Santa Monica, CA: Human Factors Society.

Crampton, J. (1992). A cognitive analysis of wayfinding expertise. *Cartographica,* **29**(3), 46-65.

Donner, K.A., McKay, T., O'Brien, K.M., and Rudisill, M. (1991). Display format and highlighting validity effects on search performance using complex visual displays. *Proceedings of the Human Factors Society 35th Annual Meeting.* Santa Monica, CA: Human Factors Society.

Drury C.G., and Clement, M.R. (1978). The effect of area, density, and number of background characters on visual search. *Human Factors,* **20**, 597-602.

Eriksen, C.W., and Webb, J.M. (1989). Shifting of attentional focus within and about a visual display. *Perception and Psychophysics,* **45**(2), 175-183.

Eley, M.G. (1988). Determining the shapes of land surfaces from topographic maps. *Ergonomics,* **31**, 355-376.

Fisher, D.L., Coury, B.G., Tengs, T.O., and Duffy, S.A. (1989). Minimizing the time to search visual displays: The role of highlighting. *Human Factors,* **31**(2), 167-182.

Fisher, D.L. and Tan, K.C. (1989). Visual displays: The highlighting paradox. *Human Factors,* **31**, 17-30.

Goldberg, J.H., MacEachren, A.M., and Korval, X.P. (1992). Mental image transformations in terrain map comparison. *Cartographica,* **29**(2), 46-59.

Haber, R.N. (1987). Why low-flying fighter planes crash: Perceptual and attentional factors in collisions with the ground. *Human Factors,* **29**(5), 519-532.

Hackworth, D.H. (1991, November 18). Killed by their comrades (friendly fire). *Newsweek,* **118**, 45-46.

Hickox, J.C., and Wickens, C.D., (1996). *Navigational checking: A model of elevation angle effects, image complexity, and feature type.* University of Illinois Institute of Aviation Technical Report (ARL-96-4/NAWC-ONR-96-1). Savoy, IL: Aviation Research Laboratory.

Kaptein, N.A., Van der Heijden, A. H. C., and Theeuwes, J. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *Journal of Experimental Psychology: Human Perception and Performance,* **21**(5), 1053-1069.

Kleiss, J.A., (1995). Visual scene properties relevant for simulating low-altitude flight: A multidemensional scaling approach. *Human Factors,* **37**(4), 711-734.

Lee, J., and Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics,* **35**(10), 1243-1270.

MacDonald, W.A., and Cole, B.L. (1988). Evaluating the role of color in a flight information cockpit display. *Ergonomics,* **31**(1), 13-37.

Madden, D.J., and Allen, P.A. (1989). Amount and duration of attentional demands during visual search. *Perception and Psychophysics,* **45**(6), 577-585.

Martens, M., and Wickens, C.D. (1995). *Lowlighting solutions to display clutter.* University of Illinois Institute of Aviation Technical Report (ARL-95-11/NASA-95-4). Savoy, IL: Aviation Research Laboratory.

Müller, H.J., Heller, D., and Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception and Psychophysics,* **57**(1), 1-17.

Nagy, A.L., and Sanchez, R.R. (1992). Chromaticity and luminance as coding dimensions in visual search. *Human Factors,* **34**(5), 601-614.

Pizlo, Z. and Salach-Golyska, M. (1995). 3-D shape perception. *Perception and Psychophysics,* **57**(5), 692-714.

Sarter N.B., and Woods, D.D. (1994). Pilot interaction with cockpit automation II: An experimental study of pilots' model and awareness of the flight management system. *The International Journal of Aviation Psychology,* **4**(1), 1-28.

Scanlan L.A. (1977). Target acquisition in realistic terrain. *Proceedings of the Human Factors Society 21st Annual Meeting.* Santa Monica, CA: Human Factors Society.

Shontz, W.D., Trumm, G.A., and Williams, L.G. (1971). Color coding for information location. *Human Factors,* **13**(3), 237-246.

Schreiber, B., Wickens, C.D., and Alton, J. (1995). *Navigational checking: The influence of 3D map rotation and scale.* University of Illinois Institute of Aviation Technical Report (ARL-95-9/NAWC-95-1). Savoy, IL: Aviation Research Laboratory.

Schulte, A., and Onken, R. (1995). *Modeling of pilot's visual behavior for low-level flight.* Synthetic Vision for Vehicle Guidance and Control. Orlando, FL: AeroSense.

Schultz, E.E. Jr. (1986). *Decluttering methods for computer-generated graphic displays.* California Intsitute of Technology Invention Report (NPO-16733/6247). Pasadena, CA: Jet Propulsion Laboratory.

Singh, I.L, Molloy, R., and Parasuraman, R. (1993). Automation-induced "complacency": Development of the complacency-potential rating scale. *The International Journal of Aviation Psychology,* **3**(2), 111-122.

Snyder, H.L. (1973). *Dynamic visual search patterns.* Committee on Vision: Division of Behavioral Sciences National Research Council. Washington, D.C.: National Academy of Sciences.

Spiker, A., Rogers, S.P., and Cincinelli, J. (1986). Selecting color codes for a computer-generated topographic map based on perception experiments and functional requirements. *Ergonomics,* **29**(11), 1313-1328.

Stiff, J. (1993). Natural scene templates as cueing devices. *Proceedings of the 7th International Symposium on Aviation Psychology.* Columbus, OH: Department of Aviation, Ohio State University.

Tan, K.C., and Fisher, D.L. (1987). Highlighting and search strategy considerations in computer-generated displays. *Proceedings of the Human Factors Society 31st Annual Meeting.* Santa Monica, CA: Human Factors Society.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lectures. *Quarterly Journal of Experimental Psychology,* **40A**(2), 201-237.

Treisman, A., and Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance,* **16**, 459-478.

Van-Orden, K.F., Divita, J., and Shim, M.J. (1993). Redundant use of luminance and flashing with shape and color as highlighting codes in symbolic displays. *Human Factors,* **32**(1), 61-77.

Wang, Q., Cavanagh, P., and Green, M. (1994). Familiarity and pop-out in visual search. *Perception and Psychophysics,* **56**(5), 495-500.

Whitaker, L.A., and Cuqlock-Knopp, G. (1991). Use of visual cues by orienteers: An analysis of interview data. *Proceedings of the Human Factors Society 35th Annual Meeting.* Santa Monica, CA: Human Factors Society.

Wickens, C.D. (1997). Frame of reference for navigation. In D. Gopher and A. Koriat (Eds.), *Attention and performance, Vol. 16.* Orlando, FL: Academic Press.

Wickens, C.D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: Harper Collins.

Wickens, C.D., Schreiber, B., and Renner, G. (1994). *3-D navigational checking: A conceptual model and an experiment.* University of Illinois Institute of Aviation Technical Report (ARL-94-9/NAWC-94-1). Savoy, IL: Aviation Research Laboratory.

Williams, H.P., Hutchinson, S., and Wickens, C.D. (1996). A comparison of methods for promoting geographic knowledge in simulated aircraft navigation. *Human Factors,* **33**, 355-376.

Wolfe, J.M. (1994). Visual search in continuous, naturalisitic stimuli. *Vision Research,* **34**(9), 1187-1195.

Yeh, M. and Wickens, C.D. (1997). *Effects of color coding, intensity coding, and decluttering on visual search in electronic map displays.* University of Illinois Technical Report. Urbana, IL: Beckman Institute for Advanced Science and Technology.

THE EFFECTS OF HIGHLIGHTING, VALIDITY, AND FEATURE TYPE ON AIR-TO-GROUND TARGET ACQUISITION PERFORMANCE

BY

RENA ADRIA CONEJO

B.S., United States Air Force Academy, 1996

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Psychology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 1997

Urbana, Illinois

THE EFFECTS OF HIGHLIGHTING VALIDITY AND FEATURE TYPE ON AIR-TO-GROUND TARGET ACQUISITION PERFORMANCE

Rena Adria Conejo, M.S.
Department of Psychology
University of Illinois at Urbana-Champaign, 1997
Dr. Christopher D. Wickens, Advisor

Air-to-ground target acquisition is an important part of any flight task, whether during navigation, landing at an airport, or during military bombing missions. This task usually involves visual searches to cross-check between a map and the forward field of view (FFOV). Previous research has implied that cultural (man-made) targets and lead-in features may yield better performance than natural targets and lead-in features. Also, since the target acquisition task often involves visual searches in complex visual fields (for both the map and the FFOV), previous research suggests that highlighting a target on the map can automate the search process by increasing the target's salience. However, as with any automation, the possibility of failure exists and is often associated with drastic consequences. This study examined the differences in performance between natural and cultural feature types as targets and lead-in features under highlighted and nonhighlighted conditions. Also, performance under highlighted and nonhighlighted conditions were compared to determine if highlighting did facilitate the target acquisition task. When targets were highlighted, the highlighting was either valid, invalid with the wrong target highlighted, or invalid with the wrong target highlighted and the correct target absent from the FFOV. The target highlighting occurred under 60% validity conditions; therefore, in some cases, a nearby lead-in feature was highlighted (always validly) to minimize or eliminate any costs of highlighting invalidity. Analysis showed that performance according to feature type was best under a target by lead-in interaction where the target feature type was opposite that of the lead-in feature type. Valid highlighting did not provide any significant benefit over nonhighlighted conditions; yet invalid target highlighting produced performance costs that were not sufficiently improved by any feature-type interaction or the lead-in highlighted condition. Further analyses suggest that performance under target absent conditions may result from different cognitive processes than those involved in conditions where the target is in the FFOV. Hence, target highlighting is not recommended unless accuracy of highlighting can be guaranteed. Also, further study must determine if the beneficial target by lead-in feature interaction reported here is generic to all tasks and environments.

TABLE OF CONTENTS

Introduction

Some aspects of aviation can be considered a target-acquisition task. Whether a military pilot who is looking for a bombing target, or a general aviation pilot who is finding an airfield located in a city, pilots must navigate to a given location and find an appropriate target when they determine they have reached the correct location. Target acquisition is one of the most important aspects of flying. No pilot can effectively plan a mission without knowing the exact details of the final destination. Also, no pilot can determine if the final destination is reached without some degree of certainty that the correct flight path has been maintained. To determine the correct flight path, the pilot must employ navigation strategies while using a preflight planning tool, such as a map, to determine the route to the destination and the features that will occur along the way.

Then, as the flight proceeds, the pilot must continually cross-check between a map, the world, and a mental model of where the target is and the features of the world along the path to the target, a process that can be described as a series of target acquisitions. Any error in this process can result in either a slight deviation from an intended path, or becoming totally lost. The errors along the route to the final target acquisition cause the pilot to miss the ultimate goal of any particular flight: the final target acquisition. Whether bombing a hostile target or landing in a distant city, acquiring the final target is the purpose for the individual flight. If this final target is missed, the entire flight may have been in vain.

Once the pilot approaches the final destination, the map serves an even more important purpose, because the map can often provide the visual cue for finding, and confirming, the final target. (We note here that the term "map" can refer to additional augmented imagery, such as satellite imagery or computer generated data bases.) Many times this target itself is an unfamiliar entity or is located in an unfamiliar background. The pilot must use the map to determine the exact location of the target. The map must serve as a discriminatory tool for detecting the correct target among numerous, and often similar, distractors. The consequences for failing to find the target, or selecting an incorrect target can be drastic (Williams, Hutchinson, and Wickens, 1996).

Modern technology has automated many of the cognitive demands imposed by maintaining the flight path. Standardized flight paths to navigational points are used throughout all genres of aviation. Also, autopilots maintain a programmed flight path to a given location. However, automation must be

monitored by the human pilot to ensure understanding of the state of the system and to enable the pilot to effectively take control of the aircraft should the automation fail. Also, established flight paths may not be available in such cases as free flight, military bombing runs, or rescue missions in unfamiliar areas. Finally, target confirmation and engagement still depend critically on the pilot visually identifying the target (Hickox and Wickens, 1996). Therefore, the pilot must engage in three general procedures: preflight route planning, navigation to a location, and detection of a particular target. The first two stages determine the pilot's visual search and expectancies during the final target acquisition. Therefore, it is important to understand how the pilot uses the information gathered in preflight planning and navigation to search the final target area, and how auxiliary maps can support the task. This study will examine possible automation-based improvements of electronic maps to ease pilots' workload in target acquisition. To understand how to improve the map, one must first consider how the pilot uses the map. The use of maps is involved in all phases of the flight task and knowledge of the flight task requirements and the associated cognitive demands will illustrate how improvements can be implemented. In the following pages, we first describe the flight task from a navigational perspective to illustrate the utility of the map during all stages of the flight mission. We then describe the visual search processes the pilot uses while scanning the map and searching the visual field for a target. Next the issue of highlighting is discussed to examine its potential benefits to the visual search paradigm used in the flight task, as well as the consequences involved when this highlighting may be incorrect, or invalid, due to unexpected failures in intelligence, equipment, etc. The following sections will demonstrate how the issues in this experiment will be useful for improving electronic maps for air-to-ground target acquisition performance.

The Flight Task

<u>Preflight preparation</u> During preflight preparation, pilots study a map to determine their navigation strategies (Schulte and Onken, 1995). In this phase, expert wayfinders usually note the general location of the target, then focus most of their attention on finding the destination within a 3 to 5 mile radius of the target for final approach (Crampton, 1992; Carmody-Bubb and Dunn, 1996). Pilots attend to salient fixpoints that bracket their course (Williams, et al., 1996; Carmody-Bubb and Dunn, 1996). Fixpoints for navigation have good visibility, are relatively unique and permanent, and can be used as <u>lead-in</u>

features. Lead-in features can be detected before the target itself is in view, but lead to the target when tracked (Schulte and Onken, 1995).

Lead-in features can be either natural or man-made (cultural). The key for their use is their salience both on the map and in the world (Schulte and Onken, 1995). Examples of lead-in features can be roads, villages, bridges, structures, and other cultural (man-made) features. Natural lead-in features might be distinctive terrain features such as ridges, lakes, water, and fields. The most important aspect of lead-in features is their contrast with the surrounding environment, either in color, absence or presence of vegetation, or vertical development (Kleiss, 1995). Lead-in features should also be distinctive compared to the surrounding environment. Pilots use these unique, salient features as "confirmation points" for the correct flight path, and may even select other cues as "back-up" cues to further confirm the correctness of the path and its lead-in features (Carmody-Bubb and Dunn, 1996). The effectiveness of the cultural versus natural features is an issue that will be discussed later.

The preflight planning process is one requiring a world-referenced display (Wickens, 1997). The pilot must determine long-distance navigational plans in relation to a wider range of space than that available from the current forward-field-of-view (FFOV). In this case a 2D, top down view best provides global awareness. In addition, the exocentric view, showing "where I am" in relation to the target provides a flight path that can be studied to arrange lead-in features into a spatially appropriate array for navigation (Eley, 1988).

This map should be oriented to the direction of travel for spatial compatibility and reduced cost of transformation, since mental rotation costs occur whenever travel is away from the canonical orientation (Aretz, 1991; Williams, et al., 1996; Wickens, 1997).

Navigational Checking The next major cognitive task after preflight planning is navigation to the destination. The wayfinder uses available information to anticipate features along the route. This includes the mental model formed in preflight planning, as well as any perceptually available navigation information, such as additional maps or cockpit images, or verbal descriptions of headings to landmarks to be encountered en route (Crampton, 1992). The navigation task is important to prevent getting "lost" enroute to a destination. This failure of a pilot or aircrew to recognize and/or maintain a desired position relative to the ground and airspace environment is defined as "geographic disorientation" (Antunano, Mohler, and Gosbee, 1989). Geographic disorientation is relatively common and can be experienced by

all types of pilots, regardless of experience or meteorological condition (Antunano, et al., 1989; Williams, et. al., 1996). The consequences of geographic disorientation can be costly in monetary terms when a pilot lands at a wrong airport (Antunano, et al., 1989). Even more significant, geographical disorientation in military bombings can result in loss of civilians' and of friendly troops' lives (Hackworth, 1991).

To prevent such geographical disorientation, pilots maintain navigational awareness through "navigational checking" (Wickens, Schreiber and Renner, 1994). In this ongoing process, the pilot compares characteristics of the map with characteristics in the FFOV to determine whether or not the two are congruent (Aretz, 1991). Navigational checking may not require an overt action if it is determined that the flight is still on the correct navigational path, but the process continuously occurs throughout the visual contact flight.

<u>The Navigational Checking Model</u> The continuous process of navigational checking can be modeled as a series of discrete matches comparing the two visual images. Hickox and Wickens (1996) developed the graphical model shown in Figure 1.1 to integrate the relevant research findings. The model represents comparison of the map stimulus (Sm) to the FFOV stimulus (Sf). Usually, viewing the map, Sm, occurs before scanning the FFOV, Sf. These comparisons are presented along a sequential time axis and represent the visual scanning that occurs in the navigational checking process. In most navigation tasks, this process occurs as subsequent iterations of Sm/Sf, without covert responses as long as the result of the checking indicates the pilot is still on the correct path (Hickox and Wickens, 1996).

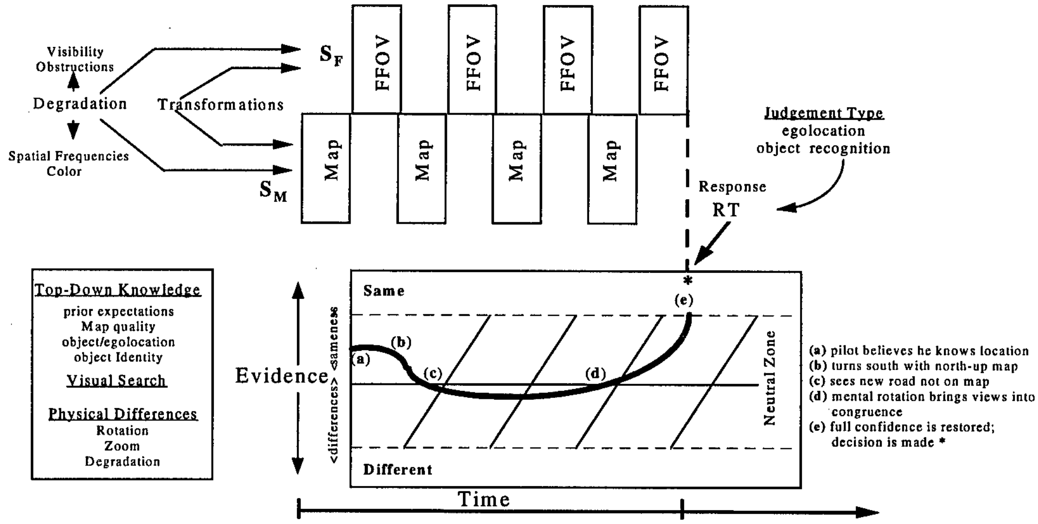


Figure 1.1: Navigational Checking Model (Hickox and Wickens, 1996)

Variables in the model that influence the efficiency of the checking process are shown on a time axis at the bottom of the figure. The model represents the subjective belief or confidence of a match between the map and FFOV on the top of the graph, and belief of mismatch between the map and FFOV on the bottom of the graph. A strong belief in a mismatch would indicate that the pilot is lost and is relatively certain that s/he is. The strength of the same/different evidence sources is represented vertically on the graph. The evidence sources are accumulated to influence the confidence that either the map and FFOV depict the same or different (i.e., lost) physical location. The various sources of evidence are accumulated over time and the NET effect of the sources, depicted by the heavy black line, determines the tendency toward a match or mismatch judgment. When the net effect crosses out of the region of uncertainty, such that the navigator is willing to commit to a decision of lost or on-track, then, if necessary, a response is made (Hickox and Wickens, 1996).

The key factors that influence the evidence accumulation and decision process may be categorized as top-down or bottom-up processing influences (Hickox and Wickens, 1996). Top-down influences on the matching processes are prior expectations that create the a priori expectancy that one is on the correct path and will therefore view a match between the map and FFOV. This expectancy establishes the initial setting of the evidence variable in a biased direction toward a "same" judgment. Top-down influences also include knowledge of map characteristics, such as errors or mismatches of the map in regard to the world. For example, the pilot knows the map may not have the same orientation,

color coding, and level of detail as the FFOV. Top-down knowledge also includes knowledge of past navigational checking successes or failures and associated behavior to maintain or correct position. Top-down knowledge may include biases from other sources of information, such as being given a specific heading toward the target, or, as we discuss below, automated map features that highlight target and/or lead-in features. It is important not to allow prior beliefs for confirmation to continue to influence subsequent sampling and interpretation of top-down perceptual evidence, in order to avoid the "confirmation bias" where evidences of match are subjectively weighted more heavily than evidence of mismatch, even when there is more evidence for a mismatch (Hickox and Wickens, 1996; Antunano et al., 1989).

By contrast, bottom-up processes draw attention to critical salient features (Hickox and Wickens, 1996). The navigator should compare the critical features of the map and FFOV conservatively, without prior expectations based on the other information sources described above. Evidence for sameness emerges when the features of the two images coincide. The salience of the features and the number of matches determines the speed with which this evidence emerges to cause the navigator to make a "same" response (Hickox and Wickens, 1996). Conversely, evidence for difference emerges when the features of the two images do not coincide. This process is more involved than evidence for sameness, since any physical differences between the two views (e.g., the map is represented from a different viewpoint from the FFOV) produce a rapid and automatic mismatch registration that must be reconciled as necessary transformations (e.g. mental rotation to align viewpoints) are performed. Also, the true differences between the map and FFOV must emerge and be aggregated as evidence for difference (Hickox and Wickens, 1996). Bottom-up processes also account for the effects of degradation of the map or FFOV visibility, transformations in scale where objects are portrayed at different sizes and separations from each other, and mental rotation between the 2D map and the 3D world (Hickox and Wickens, 1996).

The processes and variables of navigational checking illustrate the importance of the auxiliary map for navigation. It is important to ensure that both the bottom-up and top-down processes are supported by the map image. An image that is congruent with the FFOV will decrease the need for transformations and will provide salient features for emergence of same/different evidence. It is important that the map is designed to be somewhat congruent with the FFOV viewed during target acquisition, which is usually performed in low-level flight. In low-level flight, pilots note presence or absence of terrain vertical development (Kleiss, 1995). The terrain is an important cue to the pilot

regarding where s/he is in relation to the world and the target. The pilot needs information about the size and clearance of terrain objects in the flight path (Haber, 1987). In addition, terrain features are cross-checked with the mental model of where lead-in and target features are in relation to the terrain (Snyder, 1973). Most important, while searching for a lead-in feature or target, the pilot is actively scanning the terrain ahead. Aviators pay attention to different kinds of visual information when flying over different types of terrain (Carmody-Bubb and Dunn, 1996). In order to facilitate this cross checking of a map with the terrain view in the FFOV to ensure correctness of the flight path, we argue below that it is important that the map be made somewhat congruent with the FFOV (Wickens, 1997; Hickox and Wickens, 1996; Schreiber, Wickens, and Alton, 1995).

Navigational checking is most efficient when the map and the FFOV are in congruence (i.e., the same scene from identical viewpoints). That is, when the map and the FFOV present the same physical scene. Research has established that at least three aspects of this viewpoint similarity are important. First is the need for congruence between the map and FFOV azimuth rotation. For example, when objects on a map are oriented differently around the vertical axis than in the FFOV, the pilot must mentally rotate one of the views to determine if they match. This rotation may produce time and accuracy costs, especially when workload is high, as the amount of rotation differences increase, at least beyond 45 degrees (Goldberg, MacEachren, and Korval, 1992; Aretz, 1991; Schreiber et al., 1995; Wickens, 1997).

Second, congruence is also supported when the elevation angle of the map matches that of the FFOV. Vertical rotation can range from a 90 degree, 2D "God's eye" view of the map to a 0 degree forward view (Wickens, 1997). The elevation angle of any viewpoint determines how much vertical information is presented to the pilot. For example, Figure 1.2 shows some different elevation angles that a map could display to a pilot and how they match the FFOV. Research has shown there is a cost for vertical rotation, when the map and the FFOV do not match (Schreiber, et al, 1995; Hickox and Wickens, 1996; Wickens, 1997). From Figure 1.2, one can intuitively understand the costs involved in this rotation. One way to counter these costs is to find an optimal elevation angle for the map. Hickox and Wickens (1996) propose that this angle is 45 degrees, since they found that it produced the smallest costs for an elevation angle disparity between the map and the FFOV whether the FFOV elevation angle was greater or smaller than that of the map. Another method of eliminating the costs of angle disparity

is to have a map that always matches what is in the FFOV by updating its viewpoint on the basis of the plane's position.

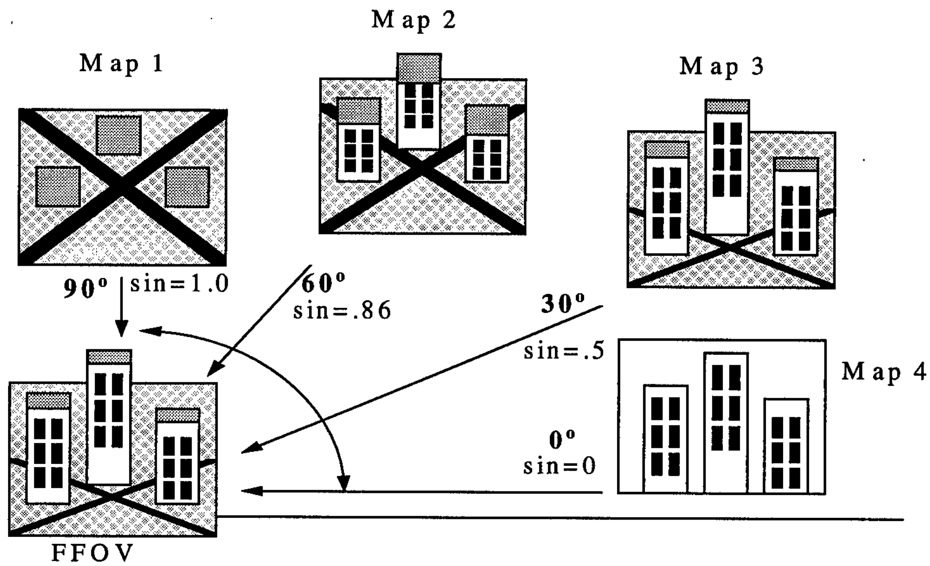


Figure 1.2: Representations of differences in feature resolution between the map and the FFOV at different elevation angles.

Third, dynamically updating maps maximize congruence because the map viewpoint matches the pilot's viewpoint through all the turns, climbs and descents. However, there are costs associated with this map feature as well. In low-level flight, the FFOV angle is very shallow, producing a "key-hole" effect with only a small amount of mostly vertical development visible (Wickens, 1997). This effect can be particularly problematic when flying toward hills and other rising terrain. For example, if the flight task is a low-level pass over a ridge and across a valley, the features of the valley may be partially or completely obscured, depending on the height of the ridge. In this case, it would still be best to keep the elevation angle higher than that of the FFOV, so the pilot can know where s/he is in relation to the rest of the world. Such a view should be accompanied by an icon used to show the placement of the aircraft in the world. The exocentric, rotating 3D map described here is best suited across all tasks of navigation, navigational checking, and spatial awareness of the features in the world (Wickens, 1997; Hickox and Wickens, 1996).

Another problem with continuously updating maps is the tremendous amount of graphics capability imposed by the continuous updating of complex 3D map images. Receptive channels in the

cockpit may not have a bandwidth wide enough to receive all this graphical information (Wickens, 1997). Again, a higher elevation angle (and more distant view of the terrain) could help this problem by reducing the amount of dynamic updating required. Fortunately, the reduced performance costs of the 45 degree map elevation angle found by Hickox and Wickens (1996) support the use of a minimally updating map or a series of static images showing continuous views of the terrain as the flight progresses.

Feature types A key element in the navigational checking task is the use of lead-in features. The lead-in features selected during the preflight planning stage facilitate the judgments about whether the pilot is on the correct route. However, there is some question about which feature types, natural (e.g., hills, water, vegetation, etc.) or cultural (man-made) features (e.g. bridges, buildings, roads, etc.), are most useful to the pilot. Both are used as cues for navigation and target-acquisition tasks, especially if they are distinct, highly-visible features (Schulte and Onken, 1995; Carmody-Bubb and Dunn, 1996).

Both types of features have similar properties which are useful for navigation, with individual advantages that may or may not make one feature type more useful for highlighting. Natural features such as mountains, hills, rivers, and others are generally highly visible and have high contrast with surrounding features. The visibility and contrast of natural features, unlike cultural (man-made) features, are usually detectable from greater distances and elevations. The presence or absence of vegetation and vertical development are features that are especially very visible and relevant during low-level flight (Kleiss, 1995). Also, natural features are more time-enduring, meaning they do not change much over time. Natural features, unlike man-made structures, generally can not be quickly or completely destroyed. Navigation training often emphasizes this enduring property of natural features and therefore suggests their use as navigation cues (Battiste and Delzell, 1991).

On the other hand, cultural features may also be highly visible and contrasted with their surroundings. While not as enduring as natural features, cultural features are also relatively permanent, barring catastrophic occurrences. Cultural features also have advantages in geometry, since their structures often have discrete, uniform, and distinct lines and angles, unlike natural features, which are usually continuous and smooth. This geometry often makes the feature view-point invariant, meaning a structure is identifiable as the same structure regardless of viewpoint (Pizlo and Salach-Golyska, 1995; Biederman and Gerhardstein, 1993). For example, one can identify several pictures taken from different

viewpoints of a house as the *same* house, but pictures of different perspectives of a mountain or a river often can not easily be identified as the same mountain or river. Correspondingly, a 90 degree intersection of two roads, or two parallel roads, appears invariant independent of its viewing perspective. Furthermore, given this invariance, cultural features can be more reliably used to judge the viewer's own viewing slant angle relative to the terrain. For example, if the viewer assumes that two roads are parallel, s/he can use the degree of convergence perceived in the roads to judge the slant angle from which they are viewed.

Some studies have tried to determine which feature type promotes the best navigation performance. Whitaker and Cuqlock-Knopp (1991) examined which cues were used most by orienteers in an off-road ground-navigation task. The cues were categorized as man-made, land contours, water, vegetation, and other. Overall, natural features (contour, water, vegetation) were used more often than man-made features. However, the frequencies of use alone do not determine which feature type is more effective. The overwhelming use of natural features could be the result of the fact that there were very few man-made features in the off-road environment used by Whitaker and Cuqlock-Knopp; and the proportion of natural or cultural cues available and used was not reported. Perhaps every cultural feature was utilized for navigation, but since so few may have been in the environment, even a relatively small proportion of natural features available and used could obscure the overall importance of the man-made features. The study does note that the participants did mention man-made objects first when asked which features were used for the ground navigation task (Whitaker and Cuqlock-Knopp, 1991).

Carmody-Bubb and Dunn (1996) carried out another exploratory study about the visual cues used in navigation-- this time in an aviation context. One objective in this study was to use eye point-of-gaze data to examine which visual features of a target area pilots considered important. Subjects flew to three cultural targets that were rated according to their contrast and visibility respective to the complexity of the surrounding environment. Target areas were considered rich in surrounding visual cues if there were highly-visible, high-contrast features nearby. Complexity was determined by the number of surrounding cues that would obscure or be mistaken for a target. (Carmody-Bubb and Dunn, 1996). The first target, a tarmac next to a paved runway near a coastline was a high-contrast, highly-visible target in a low-complexity area. The second target, a dirt airstrip in a flat field, was a low-visibility target with poor, low-complexity surrounding cues. The third target, a small row of barracks in the middle of a city, was a

low-visibility target in a rich, highly-complex environment. The characteristics of the target and the surrounding environment are useful for interpreting the data on visual cues.

For targets one and two, pilots used mostly natural cues. For target three, pilots used mostly cultural cues. In all, seven natural features were used 37 times and six natural features were used 26 times for all three targets. However, these data do not immediately lend themselves to a determination of which feature types are used most effectively for navigation. As in the Whitaker and Cuqlock-Knopp (1991) study, the differences in feature use may be a product of the target environment. Targets one and two in this study appear to be in areas where there are only a few cultural features. Target three was in a highly-cultural environment and almost all pilots consistently used the same cultural lead-ins. The Carmody-Bubb and Dunn (1996) and the Whitaker and Cuqlock-Knopp (1991) studies are both useful for determining the frequency of use of (i.e., preference for) natural and cultural features in navigation. However, the data in these studies may be largely indicative of the respective environments of the tasks. Also, while the data indicate the frequency of use, they do not indicate the effectiveness of the relative feature types.

Battiste and Delzell (1991) attempted to answer the question of effectiveness by testing the terrain features that helicopter pilots use for maintaining geographical awareness. In one portion of the study, Emergency Medical Service (EMS) helicopter pilots drew a map of their service area from memory. These pilots were instructed to include all features they used for orientation. In another portion of the study, non-EMS pilots performed a simulated navigation task using a tactical navigation chart while seated in front of a television display of the FFOV. These pilots were instructed to verbally report the features of the display and map that were used for navigation and orientation. Also, pilots were given a questionnaire to assess the utility of different types of map features. Battiste and Delzell found that pilots reported using cultural features more than natural features in familiar environments. In unfamiliar environments, however, pilots used natural features for orientation. The types of features selected for navigation were generally determined by the familiarity of the terrain. In unfamiliar areas, these cues are frequently mission-specific according to the projected flight path. From the subjective utility of feature types ratings, pilots also displayed a similar preference for cultural features. The difference in subjective utility were not great, but the average utility of cultural features was higher than that of natural features (Battiste and Delzell, 1991).

While this utility measure was largely subjective, it is supported by findings from Hickox and Wickens' (1996) navigational checking task. In this study, pilots determined if the scene represented on an electronic map was the same as that in the projected FFOV. The scenes were categorized as either containing primarily natural or primarily cultural features. Scenes with primarily cultural features yielded both faster and more accurate same/different judgments. Subjects also reported that cultural features were most helpful for making judgments about the navigation task. Natural features also suffered a cost in speed and accuracy for scenes with low-complexity and a cost in accuracy for highly complex scenes. Finally, judgments of scenes with natural features were more impaired by differences in viewing angle between the map and FFOV, and therefore considered "viewpoint dependent" (Biederman and Gerhardstein, 1993), where as judgments of scenes with cultural features were not impaired ("viewpoint independent").

In summary, the studies on feature type suggest that navigation is best facilitated by cultural features, even in cases where natural features are used more frequently (Whitaker and Cuqlock-Knopp, 1991; Carmody-Bubb and Dunn, 1996; Battiste and Delzell, 1991; Hickox and Wickens, 1996). Also, pilots appear to prefer cultural features over natural features when selecting navigational points of reference.

Target Acquisition Once the pilot approaches the final target destination, the cognitive load increases to its maximum level. Here, the pilot must not only continue to fly the aircraft, but also engage in search for the target in the terrain. During this phase, the pilot uses the information gained in the enroute planning and navigational checking phases to search for a correct target. The new task drains more of the already limited perceptual cognitive resources the pilot must allocate to different functions of flight. For example, a military pilot on a bombing run must continue a safe flight, often at very high speeds and low altitudes, and must ensure that the aircraft is maintaining correct flight functions, such as power, heading, and airspeed. The pilot may also need to fly in terrain that is rough and sometimes unfamiliar and even hostile. Therefore, the pilot must avoid terrain and other hazards and continually check the position of the aircraft in relation to these hazards. Finally, during the target acquisition, the pilot must cross-check the FFOV against the map and the mental information about the target to locate and confirm the correct target with a relatively high degree of accuracy. The extent to which the target is explicitly represented on the map by a symbol such as a mountain or buildings, or is implicitly represented by a

general location or verbal description, such as "a portable missile launcher near a river", will vary from occasion to occasion.

This process of final target acquisition must occur within only a few seconds. Because of the high speeds aircraft often fly on combat approaches, the target may only be visible for eight to ten seconds before the time of weapons release (Stiff, 1993). It is also important to note that because of the other tasks the pilot performs in this stage, the actual time available for head-down concentration on a map is further limited. Stiff (1993) estimates this available time as not much more than 1.5 to 2 seconds. Optimum use of this limited time should thus be facilitated by the auxiliary map.

Because of these numerous, highly dynamic tasks, the final target acquisition performance is extremely important to study and optimize. Unlike the preflight planning and navigational checking enroute tasks, which have been somewhat more widely studied (Stiff, 1993; Carmody-Bubb and Dunn, 1996; Eley, 1988; Schulte and Onken, 1995; Goldberg, et al., 1992; Williams, et al., 1996; Hickox and Wickens, 1996; Schreiber, et al., 1995; Wickens, et al., 1994), the final target acquisition phase of aviation is relatively unstudied in its realistic, applied domains. Studies on the component visual search features of target acquisition, that are described below, do provide useful expectations as to how the pilot might behave in final target acquisition.

Visual Search

The major difference between the target acquisition stage and the navigational checking task involved in the enroute navigation stage is the intensive dynamic visual search in the former case. This visual search is generally across a greater number of features in the map and the FFOV than in the visual search during enroute navigation. Also, unlike the enroute navigation task, this visual search is a more bottom-up oriented search. That is, whereas navigational checking is usually a confirming task with no required overt action, the target acquisition is a more active search process with a required action whether the two visual fields are congruent or not. Once the pilot reaches the target area, the matching process must search through many features, often in greater numbers and more similar to each other, with a less spatially organized mental representation than in enroute navigation. This more bottom-up oriented task is representative of classic visual search patterns. Therefore, it is important to understand these patterns to better predict the pilot's behavior and the tools that may support the search task.

Visual search is the focusing of attention on a visual display of items (visual field). It is used when the task is to locate a particular target at an uncertain position in the visual field (Drury and Clement, 1978). Examples of visual search include searching for misplaced keys in a cluttered office, scanning a grocery receipt for the price charged for an item, or a map for a particular street. One very difficult and familiar task is searching a sports arena for a friend when a meeting place had not been determined. These tasks are difficult because other objects, called distractors, compete for visual attention. In the stadium example, all the other people at the event make it difficult to locate the friend because they must all be scanned to determine if their faces are or are not the friend's face.

Although search tasks in some visual fields require examining items one by one (in series) to find a target, some features of targets can be processed preattentively. The preattentive phase of visual search is an initial "overview" of the visual field where objects and groups of objects are organized and noted for subsequent serial attention (Treisman, 1988). When a target is defined on one level by a salient feature, different from the other features in the visual field, it "pops out" from the rest of the distractors in the preattentive phase and is an initial starting point for subsequent serial search (Wang, Cavanagh, and Green, 1994). In the stadium example, when looking for a friend with red hair, the hair color is a feature of the target that differentiates it from distractors. During the preattentive phase the people in the visual field with red hair would be noted, and during serial search, attention would be focused on every person with red hair until the friend is located.

Two important characteristics of the visual search field that affect the efficiency of the parallel and serial stages of visual search are the number of similar distractors and the complexity of the background. The major effect on search time comes from the number of distractors in the visual field. The number of similar nontargets immediately surrounding the target linearly increases search times (Drury and Clement, 1978; Carter and Cahill, 1979). This relationship explains why finding the same friend is easier in a room with only a few other people than in the stadium.(Treisman and Sato, 1990). The second characteristic of the visual field that affects search time is the complexity of the field. Complexity could be considered to be related to the number of dimensions by which distractors differ from each other in the visual field (Wolfe, 1994; Carter and Cahill, 1979; Drury and Clement, 1978). Complex visual fields have items that differ in more than one dimension. In the stadium example, not only are there a large number of people (distractors) to search through, but the people are arranged in a complex background. The people are of all shapes and sizes, and wearing a number of different colors.

Some people are seated, some are moving, but all are human-shaped within a relatively finite set of widths and heights. The visual field is complex because it has a lot of items with differing dimensions that are densely arranged in the area, but the distractors are also similar to the target because generally none are extraordinarily large or bright or otherwise distinct from the friend on immediately salient levels. Therefore, a stadium search may be very time-intensive in a similar manner to an air-to-ground search involving a visual field with a large number of distractors similar to the target arranged in colorful, complex backgrounds.

In addition to the increased search times, searching complex displays demand a large amount of attentional resources. Madden and Allen (1989) tested a visual search task concurrently with a secondary task to determine the amount and duration of attentional demands for visual search. Subjects were required to find a target letter among one or three distractor where the target set was constant throughout each trial, or varied from trial to trial. The secondary task was a simple reaction time to a tone. Subjects performed the visual search as the primary task, and the tone-detection secondary task. Performance on the secondary task decreased as complexity of the display increased. This experiment indicates that more attention must be diverted to the visual search task when the visual search is more complex, thus depriving other tasks of required attention.

To date, most of the research on visual search tasks has been performed using simple, artificial stimuli. Many of the tasks involve locating an alphanumeric or symbolic target among distractors that are also alphanumeric or symbolic. In addition, most of these visual search tasks occur on blank or homogeneous backgrounds (Wolfe, 1994). Are these simplified tasks relevant to "real-world" visual searches? To answer this question, Wolfe (1994) tested the visual search models against naturalistic terrain representations assembled to replicate aerial views of the world. Wolfe had subjects search for targets characteristic of "real-world" visual search. This experiment showed that the amount of clutter in the display did indeed increase the search time for the naturalistic stimuli, as did the continuous, natural backgrounds in which the targets were embedded. Wolfe's work is very important because he found that the models based upon controlled laboratory environments can be extended to "real-world" situations. This experiment transcends the differences between basic and applied research and opens a world of possibilities for improving visual search in complex, natural tasks.

Indeed, the direction of some visual search studies is toward more naturalistic, complex stimuli involving target acquisition. Target acquisition can be described according to the Hickox and Wickens

(1996) model of navigational checking as continued iterations of two dynamic visual searches. The pilot must (1) search the map to locate lead-in features, the target, and surrounding cues to confirm the target location and continuously confirm the map view with (2) a search of the FFOV and the mental image formed during preflight planning. Snyder (1973) studied dynamic visual search patterns for air-to-ground target acquisitions. The objective of the study was to determine how to make the time-limited, dynamically changing air-to-ground search task more efficient. Snyder reports that pilots currently perform useless searches for more than 40% of the time between when the target is visible and when the pilot makes an acquisition response. In Snyder's study, a visual scene was projected on a 15 foot radius spherical screen, through which test pilots flew two prebriefed missions, each with four targets. Eye movements were recorded to determine the patterns and objects relevant to the target search task.

The results of the study provide some very useful information about how pilots search dynamic visual scenes. First, median dwell times (320msec) were not different from those recorded in static eye-movement studies, and there was no variation in dwell times as the aircraft approached the simulated target. The search patterns were not random or distributed in geometric patterns, but were concentrated near the horizon in the center of the field, and on certain types of distinct terrain such as clearings and roads. However, it is important to note that most of the targets for this experiment occurred in such locations. Therefore, it may be that fixation points are related to the expectation of target locations. If the target is expected to be in a clearing or near a river according to the nature of the target (i.e., a bridge on a river) or to preflight planning descriptions, these areas are most likely to receive more fixations. The implications of the search patterns are also supported by the pilots' reports about the features of the target used to judge its location. Pilots reported estimating the characteristics of the target and background that would be most apparent and searched for those. The detailed appearance of the target itself was used more as a check in positive identification. Essentially, the characteristics of the surrounding terrain and the associated cultural features were found to be at least as important as the target itself.

The significance of the terrain for naturalistic visual searches is further illustrated by a study by Scanlan (1977). This experiment examined the effect of low versus high background complexity (many items with many dimensions of information), target/background contrast, and display image resolution on the time to detect a tactical vehicle target in a realistic scene. Subjects used a wooden pointer to indicate the position of a target and the experimenter visually verified the correctness of the designation.

The effect of high background scene complexity was detrimental to performance. For example, the images requiring the longest detection times were those where the target was located in an area with a number of similar sized objects. Therefore, not only was the complexity of the scenes important, but so also was the placement of the target in those scenes. If the contrast between the target and the background was high, search times decreased. Also, increased image resolution decreased the required search times. These effects indicate that in realistic backgrounds, target detection is dependent on the perception of the form of the target and its location in the visual field.

The results of this study suggest that the areas to be examined and the order in which they are examined are influenced by the global descriptions of the target and the assessment of likely target locations. The search process proceeds with examinations of locations that might contain an object with some expected target characteristics to warrant further examination. This search process is influenced by the clutter features in the scene. In summary, Scanlan (1977) reports that detection of a target in even a simple realistic background is 11 times as long as search for the same target in a uniform background. With high complexity backgrounds, the factor increases to 24. Therefore, Scanlan concludes that displays must be simplified to facilitate target searches in critical, time-limited situations. However, in the navigational checking task, simplification of augmented images by deleting or lowlighting features may not be desirable because critical navigation features may no longer be visible. An alternative, which would simplify the display while leaving all features available to the pilot, is making targets (and perhaps lead-in features) more distinctive through highlighting.

Highlighting

Highlighting is simplifying a display by intensifying or uniquely coding a small set of highly distinctive, salient target features (Schultz, 1986). When an object differs from distractors in at least one salient feature, it may be said to be "highlighted" and the object will "pop-out" from the distractors in the display, calling attention to the feature, and thus enabling it to be processed in the early, parallel stage of visual processing (Müller, Heller and Ziegler, 1995; Wolfe, 1994; Eriksen and Webb, 1989; Kaptein, Van der Heijden, and Theeuwes, 1995). attention is initially attracted to highlighted targets to facilitate the search process (Tan and Fisher, 1987). Highlighted items have a search time that is independent of the number of non-highlighted distractors; but as the number of items in a particular highlighting subset increases, the search time increases to resemble a serial search (Kaptein, et al., 1995). Three kinds of

highlighting generally yield the best results for search times: luminance, flashing, and chromaticity (color).

Luminance (also called intensity) is highlighting a target by making it brighter than the surrounding distractors. Under this highlighting condition, the distractors can also be decreased in intensity (lowlighting) if desired or necessary to prevent them from interfering with the visual search process. Flashing is dynamic highlighting where the stimulus cycles on and off. The dynamic nature of flashing with its sudden onsets and offsets immediately call attention to itself (Fisher and Tan, 1989). This highlighting method is optimal for localization, but not identification of targets because it can only be processed in the "on" stage of its cycle (Fisher and Tan, 1989). Research suggests that flashing, in conjunction with another highlighting method, may be useful as a secondary highlighting method to guide attention to other targets (Van-Orden, Divita, and Shim, 1993).

Across most studies, the most effective form of highlighting has been color (Yeh and Wickens, 1997; Fisher and Tan, 1989; Nagy and Sanchez, 1992; Brown, 1991; Converse, Kozai, Batten, 1992; Spiker, Rogers, and Cincinelli, 1986; MacDonald and Cole, 1988; Tan and Fisher, 1987, Shontz, Trumm, and Williams, 1971; Christner and Ray, 1961). Color is the most effective medium for search tasks, decreasing search times by up to 70% (MacDonald and Cole, 1988). Evidence shows that color is processed in the preattentive parallel stage for cueing location toward prospective targets (Converse, et al., 1992; MacDonald and Cole, 1988). Color attracts attention in the same way as flashing, and is useful for identification purposes as well (Fisher and Tan, 1989). The target is visible at all times in the display, so the subject can begin identification of the target immediately after localization (Nagy and Sanchez, 1992). Distinct color behaves similarly to increased luminance, but may be more effective for monitors with limited intensity capabilities and dynamic displays. Also, the attention-getting qualities of color highlighting are enhanced in tasks with greater display complexity, such as navigational checking (MacDonald and Cole, 1988; Christner and Ray, 1961).

Some studies have applied different highlighting techniques to test for the same highlighting benefits in complex tasks. Martens and Wickens (1995) and Yeh and Wickens (1997) used different highlighting techniques to facilitate focused and divided attention tasks with complex displays. Both studies found a significant benefit for highlighting methods. Subjects were able to use the different displays faster and more accurately when highlighted. Yeh and Wickens (1997) found color to be slightly more effective than intensity.

Although highlighting usually benefits search tasks, there are cases where highlighting may be more of a hindrance than an aid. Specifically, highlighting can produce costs if it is invalid. Highlighting validity is defined by the probability that the highlighted object is actually the correct target (Fisher, Coury, Tengs, and Duffy, 1989). Highlighting is most valuable in complex tasks with unfamiliar stimuli, such as finding a target in a complex and unfamiliar background. Highlighting can be considered as a means of partially automating such a complex task. Attention is drawn to the highlighted option, reducing the need to scan the entire visual field. Thus, highlighting is an automation agent which facilitates the visual search to the suspected target. However, in such complex circumstances, given the costs of invalid highlighting, the automation must be very careful in preselecting the correct targets for highlighting. Because of the great similarity of objects in such complex visual fields, there is a greater opportunity for automation to select an incorrect (invalid) target. Unfortunately, as is the case with all automation, highlighting may not always be perfectly reliable because there is no fail-safe guarantee that the target which is chosen (in advance) to be highlighted will actually be the correct highlighted item during visual search.

To explore the costs of invalid highlighting, Fisher et al. (1989) studied the effect of highlighting at 25% and 75% validity. In this study, subjects searched for a specific target word in a background of distractor words. In the first experimental condition, the target word was highlighted 75% and a random distractor word was highlighted the remaining 25% of the time. In the second condition, these ratios were reversed. For both test conditions, the target could be present or absent. Fisher et al. (1989) found that response times were significantly slower for 0.25 highlighting validity than for 0.75 validity for both target present and target absent trials, although target absent trials produced significantly slower response times. The experimenters used these results to develop a model of visual search in invalid highlighting conditions. Fisher et al. (1989) propose that in a highlighted display, subjects first search the highlighted option(s) to find the target. If the target is not highlighted, subjects then search the entire display until the target is found. If the target is not found (i.e., target absent trials), subjects return to the highlighted options for a second analysis before making a target absent response.

Fisher and Tan (1989) tested the model developed by Fisher et al. (1989) in a 50% highlighting validity condition to determine if the costs of invalidity canceled the benefits of highlighting for visual search. The results of this study suggest invalidity produces costs in highlighting, since responses were never faster, and were sometimes slower when highlighting occurred. These results were inconsistent

with those of the Fisher et al (1989) study, because in the previous study, subjects were faster for highlighting than no highlighting, regardless of validity. Therefore, Fisher and Tan conducted a second experiment to determine how subjects use highlighting in visual search. In this second experiment, the same participants performed the same task, but with a highlighting condition of 100% validity. The assumption for this second experiment was that if subjects always attended first to the highlighted option, the response times for the first and second experiments should be identical (Fisher and Tan, 1989). However, the results of the second experiment showed that subjects were faster when the highlighting validity was 100%. Therefore, as highlighting validity decreases, the benefits of highlighting are lost (Fisher and Tan, 1989).

Fisher and Tan (1989) suggest that the nature of the task itself also determines how subjects search highlighted displays when the validity of the highlighting is questionable. If the visual field is complex, such as in the Fisher et al. (1989) study, subjects obtain a benefit for attending first to the highlighted option, even if the validity is low. Since a complex display requires subjects to encode more objects, the overall response time is aided by highlighting. If the display is simple, the subject must encode fewer objects. The cost of ignoring the highlighting and performing a simple serial search is relatively small, thus producing a minimal and almost unnoticeable cost for ignoring the highlighting (Fisher and Tan, 1989).

This explanation is supported by the results of a study by Donner, McKay, O'Brien, and Rudisill (1991). This study examined the effects of highlighting validity in complex alphanumeric displays. Subjects searched two Space Shuttle information displays to answer questions about the values of various display items. Subjects viewed the information displays in their current (poorly formatted) and reformatted versions. Highlighting was present on 80% of both of these displays, equally divided into valid and invalid applications. The results of this study replicated those of Fisher et al. (1989) and Fisher and Tan (1989). For the simple, reformatted, displays, the highlighted and non-highlighted conditions were not significantly different at 50% validity. For the more complex, poorly formatted displays, valid highlighting improved search performance, and the cost of invalid highlighting was negligible (Donner et al., 1989).

The results of these studies show that in complex displays, it is best to have highlighting, even if validity is relatively low. In such complex displays, search performance for highlighting is at least as good as standard, non-highlighted displays. To retain the benefits of highlighting for visual search,

validity should be at least 50% (Fisher, et al, 1989; Fisher and Tan, 1989; Donner, et al., 1991). Unfortunately, in some tasks, the cost of invalidity may go beyond the extra time required for visual search. In tasks, such a military bombing mission, invalidity could lead to false target acquisition, and ultimately, loss of innocent lives.

The issue of highlighting validity is directly relevant to the techniques that can be used to facilitate air to ground target acquisition. Prior intelligence can identify a target (or lead-in feature) and electronically enhance it on the pilot's map. The pilot would, of course, be expected to confirm that the highlighted target is the correct target in the FFOV. Unfortunately, this process is subject to failure, driven by top-down processing when the intelligence is incorrect or the system fails and an incorrect target is highlighted on the map. Because of the high workload and time pressures involved in such a mission, pilots may choose to rely on the highlighting without confirming the correctness of the target in the FFOV. Or, if the pilot does decide to confirm the target, this judgment may be biased by the fact that the item is highlighted on the map. The pilot may discount features in the FFOV that indicate the target is incorrect and/or give more credence to similarities between the incorrect target and the actual target to conclude the target is correctly highlighted, even when it is not.

This overconfidence in automation, called complacency, is a major issue in modern aviation tasks. Complacency is overconfidence in a usually reliable system, that may result in non-vigilance behavior (Singh, Molloy, and Parasuraman, 1993). Pilots' attitudes of overconfidence and overreliance on the automation are understandable, since automation is generally very reliable. As continued use of automation remains successful, pilots' trust in the automation grows until the expectation of malfunctions is so low that pilots do not monitor the automation very closely; resulting in missed detections of automation failure (Wickens, 1992). The complacent attitude itself may not always result in complacent behavior, but when combined with some critical situations, such as high workload situations, such behavior does contribute to the probability that the operator will fail to detect or even search for automation failures (Singh, et al., 1993). The results of such complacent behavior may be particularly disastrous in situations where the automation is used to support some judgment, such as engaging a target for attack.

Lee and Moray (1992) suggest that complacent behavior may stem from other factors besides overconfidence in the automation. In addition to operators' perceived reliability of automated systems, which corresponds to trust in the automation, operators may not feel sufficiently confident in their own

abilities to perform a task, and may therefore rely on automation, despite any mistrust they may feel regarding the automation. For example, in our study, it is possible that a pilot may rely on automation to select a target, even if the highlighted object does not correspond to the characteristics determined in the preflight planning stage, simply because the pilot has more confidence in the highlighting automation than his or her own memory and mental model of the flight path or target characteristics. This complacent behavior due to mistrust in personal skills and knowledge may be as disastrous as that caused by overconfidence in automation (Lee and Moray, 1992). Whatever the cause, information must be provided to the pilots to allow them to "check" the correctness of automation performance, and thus decrease the possibility for failure.

In our study, the pilot must select a target (for a landing area or to destroy) in the simulated FFOV using an electronic map with a target that is sometimes highlighted, but with less than perfect validity. An incorrect target acquisition could mean landing in the wrong, possibly dangerous area, or destroying friendly property or life. To prevent such drastic consequences, as well as the increased search times, steps must be taken to counteract the effects of highlighting invalidity in a target acquisition task. One such countermeasure could be to make another near-by, or "lead-in," feature salient for a confirmatory cross-check. Since, as discussed earlier, the background features are important for confirming the target itself (Snyder, 1973), making a background feature salient through highlighting could combat the effects of target invalidity while maintaining the benefits of highlighting for visual search. Carmody-Bubb and Dunn (1996) found pilots are more likely to find a target when they use secondary "back-up" cues. The back-up cues, or lead-in features, are important to navigation and geographical awareness, but most importantly, they are significant to the pilot as the indicators of the path to the target. The lead-in features are expected cues, since they are chosen during preflight planning and remain in the pilots' mental representations of the target locations. For this reason, the lead-in features can help indicate highlighting invalidity when the highlighted target is not along the "path" of the lead-in features. In addition, the lead-in features themselves are less susceptible to mistakes from invalid highlighting since the very basis of their choice is that they are unique, salient, and meaningfully located along the flight path. Furthermore, they are generally permanent fixtures on the ground and hence can be more reliably and accurately placed by intelligence. The properties of lead-in features can be further exploited for target confirmation if at least one is made more salient through highlighting. The next question that arises when highlighting lead-in features is *which* lead-ins (in addition to the

target) should be highlighted. Should natural features be highlighted because of their unique, time-enduring properties and because the additional saliency may overcome the costs of viewpoint dependence? Or would the preexisting preferences and benefits associated with cultural features be increased by highlighting these features? these questions will be addressed in the present experiment.

In conclusion, the research described above has revealed that navigational checking is a critical and difficult skill that is guided by both top down and bottom up processing. Bottom up effects depend on the ease of matching map features with terrain features in the FFOV, a process that will be challenging in complex or ambiguous environments. Particularly in the former case, visual search will be challenged if there are may possible similar appearing elements, Some evidence also suggests that the challenge is greater when the scene contains primarily natural features, rather than cultural ones (Hickox and Wickens, 1996). Top down influences result from prior expectations based upon target descriptions and upon a general belief that one is heading along the correct path (i.e., one is initially oriented correctly).

The proposal to highlight electronic maps can affect both bottom up and top down processes. The influence on bottom up processes is achieved by making the targets (or other highlighted items) more salient, and we have noted highlighting benefits in basic search tasks and menu searches. The influence on top down processes results to the extent that the navigator trusts that whatever automation or intelligence is selecting (and then highlighting) is the item in question. Studies of highlighting in non aviation domains have revealed the costs of highlighting when the highlighting may be invalid. In air-to-ground search, we can envision the consequences of invalid highlighting (from faulty automation or intelligence). On the one hand, it may simply slow the process of positive target confirmation as the pilot must first reject the highlighted item, and the search for the correct one (as observed in highlighted menu searches). On the other hand, and a more serious concern, it may be that complacency in the accuracy of highlighting leads pilots to ignore bottom up cues and incorrectly approach or attack the wrongly highlighted target.

Present Experiment

This study examines three issues related to target search and identification. First, we wish to replicate the conclusion reached by Hickox and Wickens (1996) that natural features require more time to identify with less accurate responses than cultural features and apply this conclusion to a dynamic flight task. Hickox and Wickens examined this issue with static scenes. This study will compare the natural/cultural effects in a dynamic target acquisition task where both targets and lead-in features may be either cultural or natural in a 2x2 design.

Second, we examine whether the target acquisition task can be facilitated by highlighting the target on the map. Since the highlighting should "automate" the search for the target on the map and cue the pilot to a particular location in the FFOV, we expect highlighting targets will result in faster target identification and selection times, as well as increasing the accuracy of the acquisition responses compared to conditions where highlighting is not used on the map. Also, if the limitations of natural features are observed, we will examine the extent to which these limitations will be compensated by valid highlighting. We will determine whether the highlighting result in performance that is equally good for both natural and cultural features, or if natural features will still be inferior to highlighted (or even nonhighlighted) cultural features for target acquisition and navigation.

Third, we examine the consequences of invalid target highlighting. We wish to determine if the highlighting results in leading the pilot "down the garden path" of complacency such that the pilot does not visually confirm that the highlighted option is correct before designating the object as a target. Since complacency may result from either overtrust in the automation or mistrust in the pilots' personal skills (Lee and Moray, 1992), we will use a confidence measure in addition to accuracy to describe target acquisition performance. If accuracy is low, but confidence is high, the "garden path" behavior likely stems from overconfidence in the automation. On the other hand, if both accuracy and confidence are low, the overreliance on automation likely stems from a mistrust of personal skills, knowledge, etc. If pilots do exhibit complacent behavior when a target is highlighted, we examine whether valid highlighting of a nearby lead-in feature will provide an "anchor" for confirmatory navigational checking to help pilots detect when the target that is highlighted is not the correct target identified in preflight planning and thus, eliminate the costs of invalidity. We expect some cost for invalid conditions where only the target is highlighted, but validly highlighting a lead-in feature should provide bottom-up information to prevent a significant cost for invalid trials.

To investigate these issues, pilots will fly a series of approaches to targets in a simulated geographical area rendered with an Evans and Sutherland image generator. These targets will be verbally described at the beginning of each trial, and the description will be accompanied by a large scale electronic map showing the starting location and upon which the pilot may locate the target. Pilots will then fly toward the target, using a 3D electronic map as a reference; verbally confirming the target when they see it in the visual world. Under some conditions, the map will be unaugmented. Under others, the target will be highlighted and under still others, the target and the last lead-in features will be highlighted. Either the targets or the (highlighted) lead-in features will be cultural or natural in a 2X2 design. To examine the validity issue, on some (40%) of the trials, an incorrect target will be highlighted (i.e., a nearby element that is similar to the target in its appearance and description). Two types of invalid trials will be created. In one case, the original target remains present; and in the other, the originally specified target is missing. Both of these situations could plausibly result from faulty intelligence in operational settings. We hypothesize that in both cases, top down processing will lead to a tendency to designate the incorrect (invalidly highlighted) target. However, this tendency will be amplified if the original target is missing from the scene, since there is no bottom up evidence to support the correct response.

Method

Apparatus and materials

An Evans & Sutherland SPX500 generated the forward-field-of-view (FFOV) and projected it onto a 7' by 10' projection screen. A Silicon Graphics IRIS computer with a sixteen inch diagonal monitor presented the preflight information, the electronic map, and recorded first response times, confirmation time (between first response and confirmed response), and the score per trial. Subjects sat in a chair approximately 32" from the IRIS computer screen (approximately 26 degrees visual angel to the center of the screen) and 11 feet from the projector screen.

Subjects

Eighteen subjects were recruited from two different instrument flight classes taught at the Institute of Aviation at the University of Illinois. All subjects had only a private pilot rating with a total experience of at least 60 hours. All subjects received the same instructions and were paid $5 per hour for their participation. In addition, the top three performers (in terms of score, response times, and compliance with the 400' AGL flight path) were paid a bonus of $10 each.

The task

At the beginning of each trial, pilots viewed a large scale 2D preflight map. The initial 2D map displayed the starting point, indicated by an arrow point in the direction of flight, and a verbal description of the target. Pilots studied the map, which was displayed for ten seconds, noting target location and lead-in features, until the trial automatically began. The active flight was then supported with a 3D, exocentric moving map. The map perspective was a constant 45 degree elevation angle with a tether length of 15000 feet. Each pilot flew 20 approaches to targets. Each approach was approximately 3 minutes in duration. As they flew, the pilots would cross check the lead-in features displayed in the FFOV with those represented on the 3D map. Airspeed for each flight was fixed at 150 knots and altitude was flown at a commanded 400 feet above ground level. Speed was held constant by the program and was not alterable by the participant; however, the pilot could control heading and altitude to avoid traffic and terrain. Visibility in the FFOV was limited to 5 miles. Subjects verbally indicated when they thought they had the target in view by saying "target," and when they were certain

the target was or was not in the world by saying "fire" or "abort" respectively. If the target was not already in the center of a projected black sighting square in the center of the display screen, subjects then used the joystick to place the target in the square and stated their confidence (high or low) that the object in the center of the screen was indeed the correct target, or that the target was not in the world.

Experimental design

The scenes that were presented to the subjects varied on three parameters in a completely within subjects design: 1) Lead-in feature type, 2) Target feature type, and 3) The presence or absence of highlighting. Each of these variables is detailed below:

1) Lead-in feature type - Flight paths to targets were chosen according to two types of salient features enroute for lead-in fixpoints. While flying to the described target, the subjects flew a path over either primarily natural (i.e., rivers, hills, etc.) or primarily cultural (i.e., roads, structures, etc.) features. Thus, when a lead-in feature was highlighted, on half the trials it was a cultural feature and on half it was a natural feature.

2) Target type - targets were either natural or cultural features.

3) Highlighting condition- The maps for each trial were defined by three highlighting types. The first condition (nohili) did not have any highlighting of the target or lead-in features. The second highlighting type (targethili) had the target highlighted in a 2x intensity red highlight in relation to the rest of the display. The third condition (targ/leadhili) had the same style of target highlighting, but the final lead-in feature (natural or cultural) was blinking. In eight of the trials, the highlighting of the target was invalid. Either the target was present in the FFOV, but an incorrect item (a nearby similar foil) was highlighted on the map, or the target was not present in the FFOV and a nearby incorrect item was highlighted on the map. For example, if the target of an invalid trial was a particular hill, (specified in the preflight verbal description) a nearby hill would be highlighted and the actual target hill may or may not appear in the FFOV. In control (no highlighting) trials, the target was always present. Pilots were aware from the instructions that some trials may contain either of the two types of

invalidity, but the number of invalid trials was not disclosed. Each subject experienced all three types of highlighting in random order. The 20 flight approaches were presented in all three highlighting conditions between subjects. Table 2.1 shows the distribution of conditions for all trials and all subject groups. Each pilot flew through all 20 flight legs. The order of trials and highlighting condition for each approach was randomized.

Highlighting for natural targets by trial number (1-20) and group (A=subjects 1-6; B=subjects 7-12; C=subjects 13-18). Example: 4C=trial four, subjects 13-18.

Natural lead-in features

control (nohili)	**target (targethili)**	**target & lead-in (targ/leadhili)**	
2A, 4A, 5A, 19B, 9C	4B, 5B	4C, 5C	**Correct target highlighted**
	9A	9B	**Incorrect target highlighted**
	2B, 19C	2C, 19A	**Incorrect target highlighted and target absent from FFOV**

Cultural lead-in features

control (nohili)	**target (targethili)**	**target & lead-in (targ/leadhili)**	
1A, 3A, 6A, 7C, 8A	1B, 3B	1C, 3C	**Correct target highlighted**
	6B	6C	**Incorrect target highlighted**
	7A, 8C	7B, 8B	**Incorrect target highlighted and target absent from FFOV**

Highlighting for cultural targets by trial number (1-20) and group (A=subjects 1-6; B=subjects 7-12; C=subjects 13-18). Example: 4C=trial four, subjects 13-18.

Natural lead-in features

control (nohili)	**target (targethili)**	**target & lead-in (targ/leadhili)**	
20A, 13B, 15B, 17B, 18B	13C, 15C	13A, 15A	**Correct target highlighted**
	20C	20B	**Incorrect target highlighted**
	17C, 18C	17A, 18A	**Incorrect target highlighted and target absent from FFOV**

Cultural lead-in features

control (nohili)	**target (targethili)**	**target & lead-in (targ/leadhili)**	
14B, 16B, 10C, 11C, 12C	10A, 16C	16A, 10B	**Correct target highlighted**
	12A	12B	**Incorrect target highlighted**
	11A, 14C	14A, 11B	**Incorrect target highlighted and target absent from FFOV**

Table 2.1: The distribution of conditions for all trials and all subject groups.

Procedure

Each subject participated in one session lasting approximately one hour. The subjects read the instructions, then were situated in the chair and verbally reminded about the important features of the experiment. The subjects were instructed to say "target" as soon as they thought the target was or was not in view. The experimenter then pressed a key to indicate the time the subject first saw the target. Subjects then confirmed the target by saying "fire" when they were sure the target was present or "abort" when they were sure the target was absent. Again, the experimenter pressed a key to record the time between the first possible sighting of the target and the final target acquisition. Subjects then positioned the target in the center of the specified projected square area of the screen (if the target was not already in this position) and indicated their confidence (high or low) that the target they had selected was correct. The experimenter scored each trial according to the accuracy and confidence of target selection. Table 2.2 shows the scoring matrix used in this experiment. No feedback was given to the subjects concerning the accuracy of their responses. After a score was entered, the trial ended and a new trial began. Subjects were instructed to work quickly, yet accurately to avoid a speed accuracy trade-off. In addition, subjects were offered an incentive for high accuracy. Subjects received three practice trials to familiarize them with the three different highlighting conditions. A short break was offered between each of the three trial blocks; however, the participant could choose not to take the break and continue with the experiment. When subjects completed all 20 trials, they completed a post-experiment questionnaire.

	Correct target selected	**Incorrect target selected**
high confidence in choice	3	0
low confidence in choice	2	1

Table 2.2: Scoring matrix for trials based on correctness of target selection and subjective confidence in choice.

Results

All statistical analyses were performed using SPSS version 6.1 for windows. The twenty subjects provided 360 trials. Of these, all trials that exceeded three standard deviations from the population means of reaction times and lag times were removed. Also, due to computer anomalies, one trial was removed for all subjects; resulting in a total of 326 trials used for analysis. The statistical analyses were carried out in two parts: (1) an analysis of valid trials to compare whether displays with correctly highlighted targets and lead-ins provided any benefit over displays with nonhighlighted targets and how this benefit might be moderated by target and lead-in feature types (i.e., natural vs. cultural); and (2) an analysis of target and lead-in highlighted trials to determine if any costs of invalid highlighting (i.e., due to incorrect intelligence) exist for this task. Both analyses were conducted using factorial ANOVA models for each independent variable.

Valid trials analysis

This portion of the data analysis focuses entirely on valid trials in which there was either no highlighting or the highlighting correctly identified the target (combined across both target highlighting and target/lead-in highlighting). Table 3.1 shows the relevant cells from Table 2.1 that were used for this analysis. This analysis also included all conditions of target and lead-in feature types (i.e., cultural vs. natural).

control (nohili)	target (targethili)	target & lead-in (targ/leadhili)	
			Correct target highlighted

Table 3.1: Design section examined in valid trials analysis

<u>Score</u> A three-way analysis of variance (natural/cultural target X natural/cultural lead-in X highlighting present/absent) using score as a dependent variable was performed to determine if highlighting targets improves performance. This analysis of variance is collapsed across both valid highlighting conditions (i.e., target highlighted and target/lead-in highlighted). A subject's score reflects both the accuracy of the target selection and the pilot's subjective confidence in the target s/he had chosen. The overall average score was 2.46, with nonhighlighted trials averaging a score of 2.32 and highlighted trials

averaging a score of 2.63. Table 3.2 shows the analysis of variance table for this dependent variable for valid trials.

Source	SS	DF	MS	F	Sig of F
Lead-in	0.03	1	0.03	0.03	0.865
Target	4.90	1	4.90	4.69	**0.032**
Highlighting	2.34	1	2.34	2.24	0.136
Lead-in X Highlighting	0.10	1	0.10	0.09	0.758
Target X Highlighting	0.78	1	0.78	0.75	0.388
Target X Lead-in	2.14	1	2.14	2.05	0.154
Lead-in X Target X Highlighting	0.45	1	0.45	0.43	0.514
Error	203.47	195	1.04		

Table 3.2: Analysis of Variance (Score)

Results of this ANOVA revealed no significant effect of lead-in feature type, but did demonstrate a significant effect of target type on score [F(1,195)=4.69, p=0.032]. Specifically, subjects scored higher when the target was a cultural (man-made) feature (mean=2.57) rather than a natural feature (mean=2.35). Further analysis revealed this effect on score was reflected by an increase in confidence, but not in accuracy. The analysis revealed a very weak trend for a highlighting benefit. However, an additional analysis collapsed across feature types [F(1,201)=2.82, p=0.095] revealed a slightly stronger effect suggesting some benefit of highlighting for score. This trend did not occur in an analysis of the accuracy alone, without subjective confidence influences (F(1,201)=1.835, p=0.177], which suggests that the weak effect of highlighting on score is a reflection of the pilots' increasing confidence in their correct answers, rather than increasing accuracy. Finally, the analysis of variance did not reveal any significant interaction effects on score for valid trials.

Initial response time The initial response time variable is a measurement of whether highlighting a target on a map enables pilots to detect the target sooner than if the target is not highlighted. Overall initial response times averaged 50.21 seconds, with nonhighlighted trials averaging 49.45 seconds and highlighted trials averaging 51.18 seconds. Table 3.3 shows the three-way analysis of variance

(natural/cultural target X natural/cultural lead-in X highlighting present/absent) for initial response time on valid trials.

Source	SS	DF	MS	F	Sig of F
Lead-in	2635.79	1	2635.79	8.36	**0.004**
Target	150.10	1	150.10	0.48	0.491
Highlighting	139.20	1	139.20	0.44	0.507
Lead-in X Highlighting	973.51	1	973.51	3.09	**0.081**
Target X Highlighting	715.60	1	715.60	2.27	0.134
Target X Lead-in	2427.17	1	2427.17	7.70	**0.006**
Lead-in X Target X Highlighting	68.99	1	68.99	0.22	0.641
Error	61504.95	195	315.41		

Table 3.3: Analysis of Variance (Initial response time)

The analysis of variance reveals a significant benefit for cultural lead-in features (mean=47.26) over natural lead-in features (mean=52.92) . However, this effect is only interpretable by examining its marginally significant interaction with highlighting [F(1,195)=3.09, p=0.08], shown in Figure 3.1, since no difference between the two feature types exists when highlighting is not present on the map.

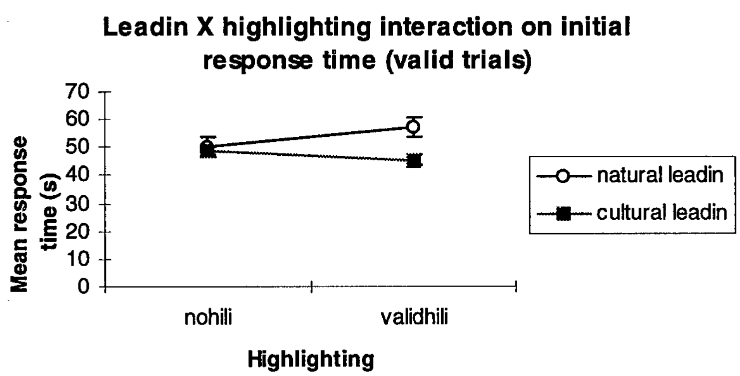


Figure 3.1: Highlighting X Lead-in interaction

The highlighting by lead-in interaction shows that a benefit exists for cultural lead-in features under valid highlighting conditions, while natural features may not be helped by valid highlighting. The benefit

for cultural features possibly stems from their smaller sizes and increased confusability with nearby cultural landmarks in their nonhighlighted state, whereas natural feature are typically larger and thus more discriminable from a greater distance in an aircraft's FFOV.

The analysis of variance for initial response time also revealed an interaction of lead-in by target type [F(1,195)=7.70, p=0.006], which is shown in Figure 3.2.

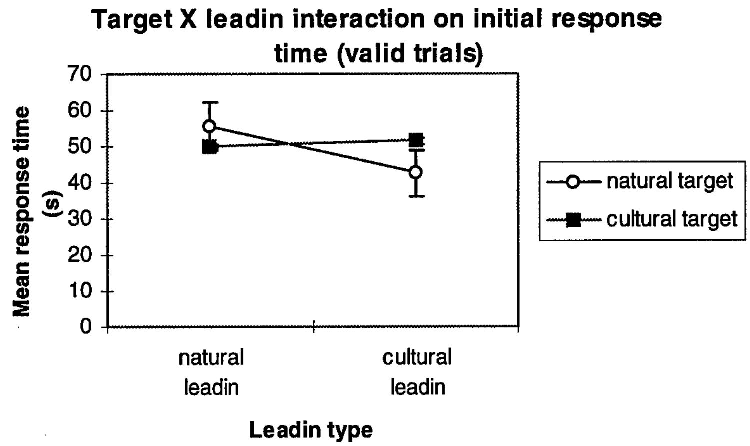


Figure 3.2: Target X Lead-in interaction (initial response time)

The target by lead-in interaction suggests that lead-in features are most helpful when they are the opposite feature type of the target. For example, cultural lead-in features help the pilot identify natural targets more quickly than cultural targets, possibly because the unique location of cultural lead-in features helps identify the natural feature in relation to the rest of the FFOV. To illustrate, when a pilot is searching for a particular hill, a cultural lead-in feature, such as an antenna, helps the pilot locate the correct hill from nearby hills much more rapidly than a natural lead-in feature, which may be confused with surrounding terrain. In addition, a natural lead-in feature reduces initial response time for cultural targets, as compared to natural targets, by providing a global awareness of where the cultural target is in the environment. The combination of global awareness and discrete location information may explain why performance is better when both feature types are present, rather than when the target and lead-in are both the same feature type.

<u>Confirmation time</u> The confirmation time is a measure of the time between when the pilot first indicates a target is in sight or absent and when the pilot is confident that the target s/he has selected is

present or that it is absent. Overall confirmation times averaged 1.52 seconds, with nonhighlighted trials averaging 1.46 seconds and highlighted trials averaging 1.59 seconds. Table 3.4 shows the three-way analysis of variance (natural/cultural target X natural/cultural lead-in X highlighting present/absent) table for confirmation time.

Source	SS	DF	MS	F	Sig of F
Lead-in	21.48	1	21.48	3.80	**0.053**
Target	6.76	1	6.76	1.20	0.275
Highlighting	0.55	1	0.55	0.10	0.756
Lead-in X Highlighting	0.07	1	0.07	0.01	0.913
Target X Highlighting	15.59	1	15.59	2.76	0.098
Target X Lead-in	1.87	1	1.87	0.33	0.566
Lead-in X Target X Highlighting		1			0.514
Error	1101.01	195	5.65		

Table 3.4: Analysis of Variance (Confirmation time)

In this analysis, the significant effect of lead-in feature type reversed that observed for initial response time. That is, confirmation time for targets with cultural lead-in features (mean=1.89 seconds) was longer than that for targets with natural features lead-ins (mean=1.18 seconds). This effect may be related to the effect of lead-in features on initial response time. Since pilots initially responded more slowly to targets with natural lead-in features, and were therefore much closer to the target at initial response, the time for confirming targets with natural lead-in features was much shorter. Essentially, in this case the pilots' initial responses and final confirmation occurred simultaneously. It is important to remember, however, that the nature of the lead-in features had no effect on score, and therefore, the differences in initial response times and lag times for natural and cultural features reflect no cost for either feature type when highlighting is valid.

The analysis of variance revealed a weak interaction (p=0.098) of target type by highlighting, shown in Figure 3.3, that is similar to the significant interaction of lead-in by highlighting on initial response time. Highlighting offers a benefit to cultural targets that is not observed for natural targets. No other significant main effects or interactions were found for confirmation time in the analysis of variance.

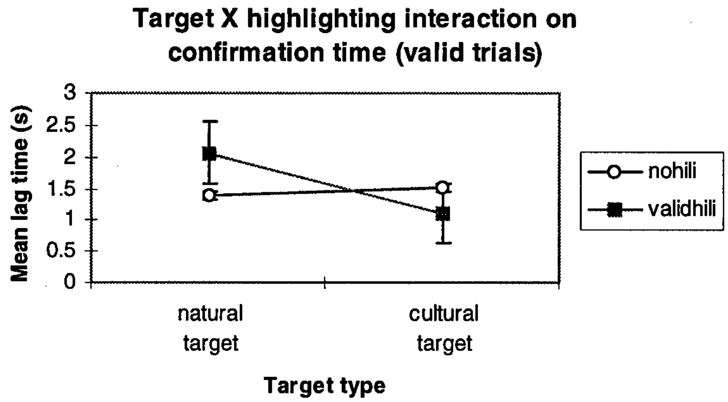


Figure 3.3: Target X highlighting interaction (Confirmation time)

Highlighted trials analysis

The previous analysis focused entirely on valid trials comparing the nonhighlighted conditions to a combination of both conditions of highlighting when the target was correctly identified by the highlighting. The second portion of the data analysis focuses entirely on highlighted trials and compared the levels of highlighting validity to determine if invalid highlighting produced a significant cost. The two types of invalidity are foreseeable examples of possible automation failure (due to poor intelligence, etc.) where the wrong target may be highlighted and the correct target is present in the FFOV, or the wrong target is highlighted and the correct target does not exist in the FFOV. Table 3.5 identifies the cells from the design table, Table 2.1, that are applicable to the highlighted trials analysis, again including all combinations of cultural and natural features.

target (targethili)	target & lead-in (targ/leadhili)	
		Correct target highlighted
		Incorrect target highlighted
		Incorrect target highlighted and correct target absent from FFOV

Table 3.5: Design section examined in highlighted trials analysis

This portion of the data analysis is where the "garden path" effect of automation complacency (i.e., failure to confirm accuracy of automation due to overconfidence in the automation) could be revealed and where the effectiveness of efforts to combat such complacency by validly highlighting lead-in features can be determined.

<u>Score</u> A four-way analysis of variance (natural/cultural target X natural/cultural lead-in X highlighting target/target&leadin X target validity: valid/wrong/wrong&notarget) was carried out with score as the dependent variable. Table 3.6 shows the analysis of variance results for the scores of the highlighted trials. This analysis reveals a highly significant main effect of validity on score, as shown in Figure 3.4. Score decreased whenever highlighting was invalid, while there was not a significant difference in score between the two kinds of invalidity, although score was slightly higher when the target was gone. Further analysis revealed that this decrease reflects a loss of accuracy in target acquisition (dashed line of Figure 3.4) without a significant loss of confidence in the pilots' selections of their targets (thick line of Figure 3.4).

Source	SS	DF	MS	F	Sig of F
Validity	56.87	2	28.43	22.32	**<u>0.000</u>**
Lead-in	8.41	1	8.41	6.60	**<u>0.011</u>**
Target	1.26	1	1.26	0.99	0.321
Highlighting	0.53	1	0.53	0.42	0.519
Validity X Lead-in	6.87	2	3.43	2.70	0.070
Validity X Target	11.91	2	5.96	4.68	**<u>0.010</u>**
Validity X Highlighting	0.10	2	0.05	0.04	0.960
Target X Highlighting	0.26	1	0.26	0.20	0.654
Lead-in X Highlighting	3.44	1	3.44	2.70	0.102
Target X Lead-in	5.14	1	5.14	4.03	**<u>0.046</u>**
Validity X Highlight X Target	0.51	2	0.25	0.20	0.107
Validity X Highlight X Lead-in	10.68	2	5.34	4.19	**<u>0.017</u>**
Validity X Target X Lead-in	2.52	1	2.52	1.98	0.161
Highlight X Target X Lead-in	3.35	1	3.35	2.63	0.107
Validity X Highlight X Target X Lead-in	2.22	1	2.22	1.74	0.188
Error	245.91	193	1.27		

Table 3.6: Analysis of variance (score)

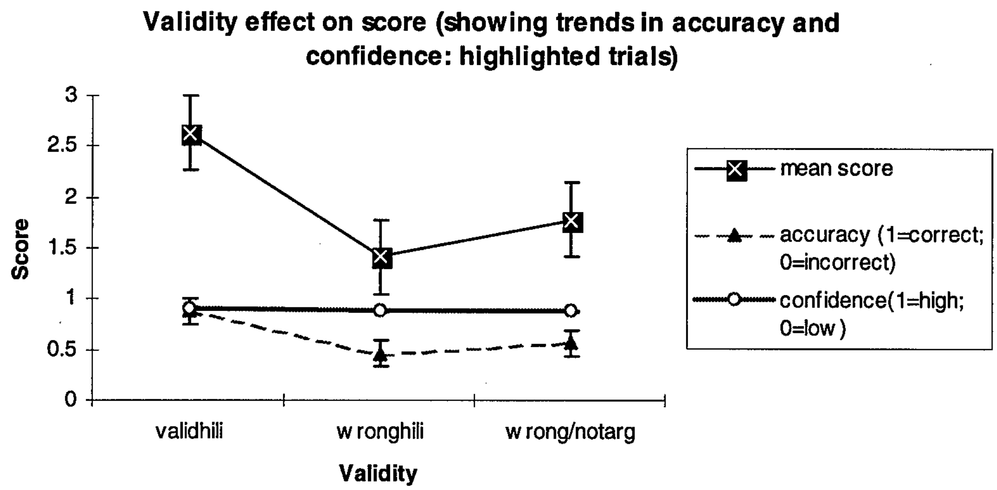


Figure 3.4: Effect of validity on score

The cost of invalid trials on score suggests that the pilots did fall victim to complacency when the target highlighting was invalid, relying on the automation, even in the few cases where they were unsure of the accuracy of their responses.

The analysis of variance also suggests an effect of lead-in feature type on score. The effect of lead-in feature type is best interpreted by examining its interactions. The marginally significant validity by lead-in interaction [$F(2,193)=2.70$, $p=0.07$] shown in Figure 3.5 suggests that lead-in feature type has no influence when target highlighting is valid, but that natural lead-in features were not as helpful as cultural lead-in features when highlighting is invalid.

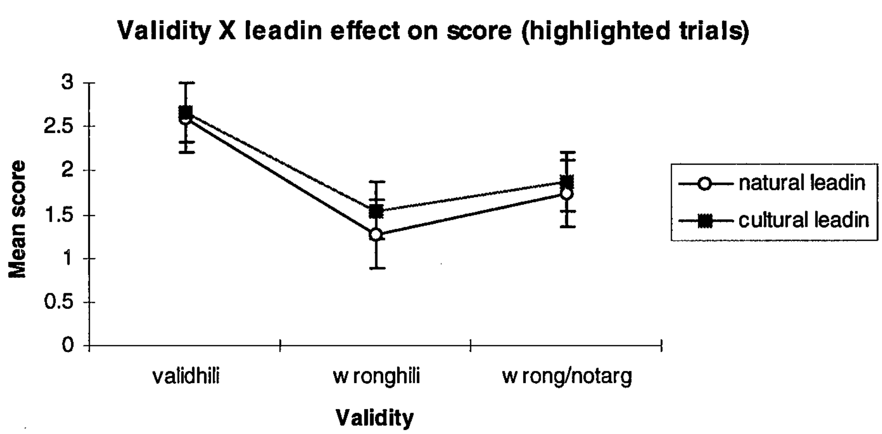


Figure 3.5: Validity X Lead-in interaction (score)

The ANOVA also shows a significant validity by target type interaction. This interaction is shown in Figure 3.6. The valid trials analysis presented above had discussed the advantage of highlighting a less salient cultural target (relative to the more salient natural target) when the highlighting was valid. In this highlighted trials analysis, Figure 3.6 reveals that the advantage becomes a cost when the highlighting is invalid, although it is not clear why the cost is diminished when the correct target is not present.

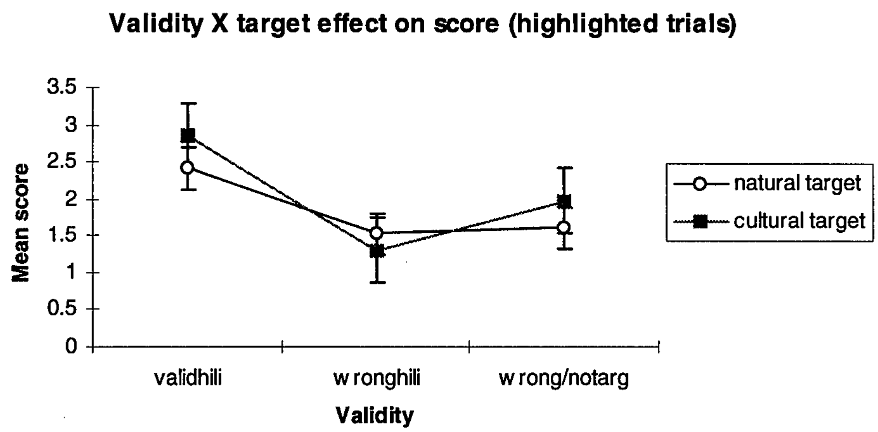


Figure 3.6: Validity X Target effect (score)

The significant target feature types by lead-in feature types interaction [F(1,193)=4.03, p=0.046] is shown in Figure 3.7 (a similar pattern of interaction to that seen in Figure 3.2), and suggests that subjects scored lower on natural targets when the lead-in features were natural and lower for cultural targets when the lead-in feature types were cultural. As we discussed above, one possible explanation for this interaction is that pilots perform better when the interaction of target and lead-in features contain both feature types; natural features to provide a global awareness of location and cultural features to provide exact location information.

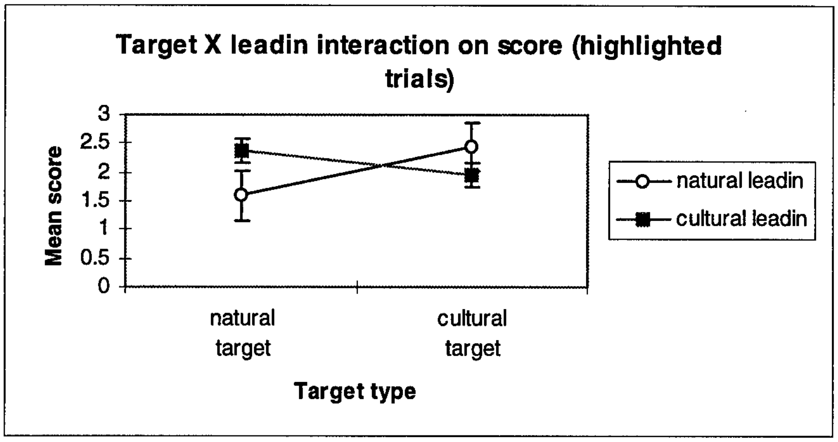


Figure 3.7: Target X Lead-in (score)

The significant validity by highlighting by lead-in type interaction [F(2,193)=4.19, p=0.017] is shown in Figure 3.8. This interaction depicts a cost for both natural and cultural lead-in feature types (left and right panels) when the target highlighting is invalid. The absence of a significant highlighting main effect or a highlighting X validity interaction suggest that highlighting lead-in features (always valid) generally did not counteract the costs of invalid target highlighting. For reasons that are unclear, Figure 3.8 suggests that the only times lead-in highlighting was successful in this role was when the lead-in feature was cultural (right panel) and the correct (prebriefed) target was missing in the FFOV (right-most points within the panel).

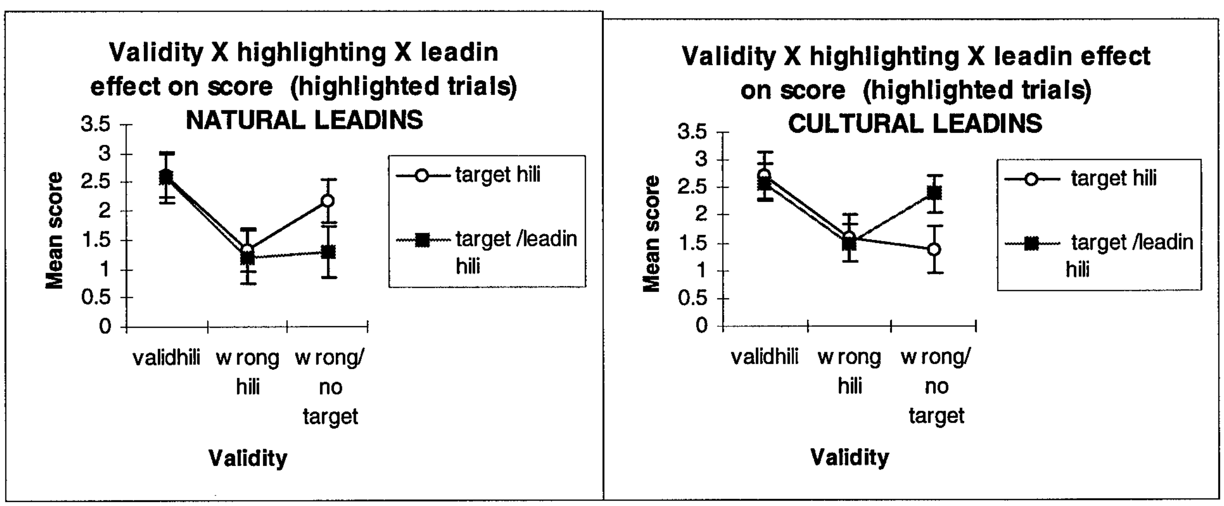


Figure 3.8: Validity X Highlighting X Lead-in interaction (score)

The ANOVA for score did not reveal significant main effects of highlighting and target type, suggesting that these variables without the additional effects of another variable in an interaction do not

influence score across conditions. Also, no significant interactions were found other than those discussed above.

Initial response time Table 3.7 shows the four-way analysis of variance for initial response time in the highlighted conditions. This analysis shows a significant effect [F(1,193)=6.69, p= 0.01] of lead-in feature type in which subjects responded faster when lead-in features were cultural features (mean=48.57 seconds) rather than natural features (mean=51.87 seconds). However, this effect is best understood by examining the very significant validity by lead-in feature type interaction [F(2,193)=6.72, p=0.002] shown in Figure 3.8.

Source	SS	DF	MS	F	Sig of F
Validity	1061.79	2	530.89	1.90	0.153
Lead-in	1872.28	1	1872.28	6.69	**0.010**
Target	945.60	1	945.60	3.38	0.068
Highlighting	123.61	1	123.61	0.44	0.507
Validity X Lead-in	3765.11	2	1882.56	6.72	**0.002**
Validity X Target	1291.31	2	645.68	2.31	0.102
Validity X Highlighting	632.89	2	316.45	1.13	0.325
Target X Highlighting	1.34	1	1.34	0.00	0.945
Lead-in X Highlighting	6.48	1	6.48	0.02	0.879
Target X Lead-in	25.20	1	25.20	0.09	0.764
Validity X Highlight X Target	1530.54	2	7675.27	2.73	0.068
Validity X Highlight X Lead-in	407.33	2	203.67	0.73	0.484
Validity X Target X Lead-in	2177.26	1	2177.26	7.78	**0.006**
Highlight X Target X Lead-in	215.23	1	215.23	0.77	0.382
Validity X Highlight X Target X Lead-in	238.87	1	238.87	0.85	0.357
Error	54029.65	193	279.95		

Table 3.7: Analysis of variance (initial response time)

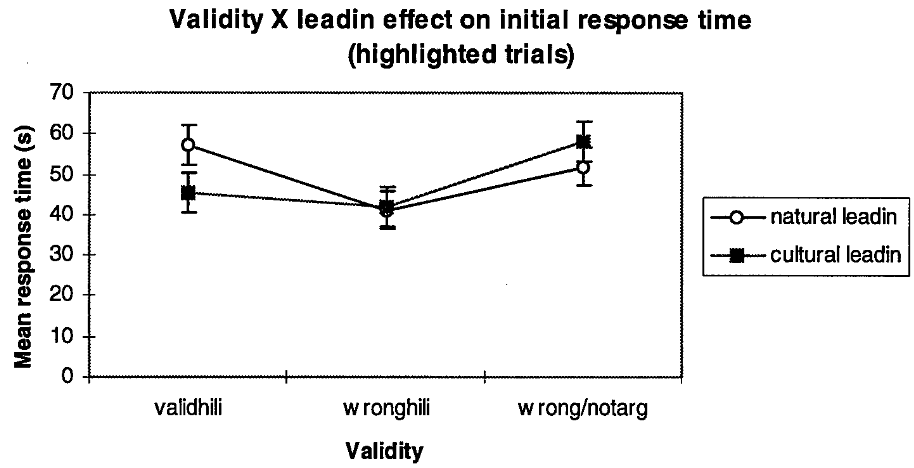


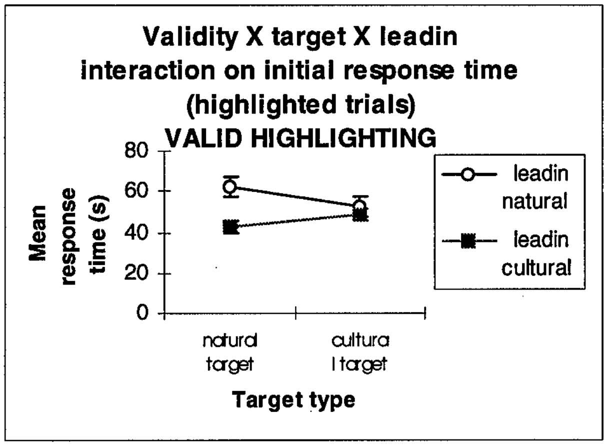
Figure 3.9: Validity X Lead-in interaction

The validity by lead-in interaction shows that subjects responded sooner when the lead-ins are cultural instead of natural, only when target highlighting is valid (the effect also portrayed in Figure 3.1). Under incorrect highlighting conditions in which the target is present (center points in Figure 3.9), initial response times to the lead-in features do not differ significantly from each other. However, when the wrong target is highlighted and the correct target does not exist in the FFOV (right points) the advantage for cultural lead-ins is reversed, becoming a cost for reasons that are unclear.

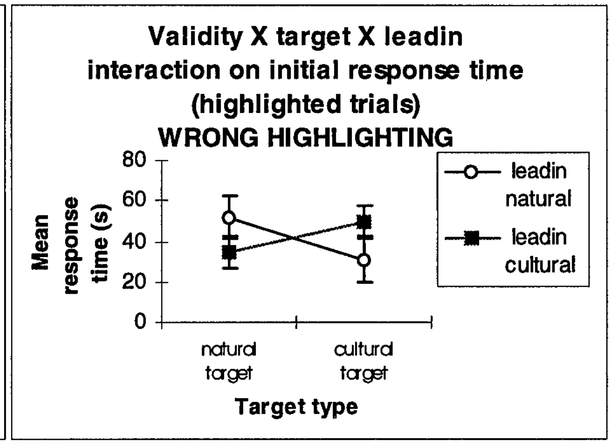
Figure 3.10 shows the three-way validity by target feature type by lead-in feature type interaction [$F(1,193)=7.78$, $p=0.006$]. This interaction shows that when the target is present in the FFOV, there is a significant benefit for cultural lead-in features over natural lead-ins when the target is natural and for natural lead-ins over cultural lead-ins when the target is cultural (Panels A and B). This interaction also demonstrates the benefit for opposite lead-in and target feature types shown in Figure 3.2 and Figure 3.6 where performance is better when the lead-in feature type is opposite that of the target feature type. However, this interaction is reversed when the correct target is missing from the FFOV (Panel C). In this target absent condition, subjects respond faster when the target type and lead-in feature type are homogeneous (i.e., natural target with natural lead-ins or cultural target with cultural lead-ins). It is unclear why the relationship for target and lead-in feature types is reversed in target absent trials from that demonstrated in the trials where the target is present.

No other significant main effects or interaction effects were found for initial response time, with the exception of a marginal effect for target feature type [F(1,193)=3.38, p=0.068] indicating fastest responses to natural targets (mean=49.6) as compared to cultural targets (mean=51.17), which is likely because natural features were generally larger and thus visible from greater distances.

PANEL A



PANEL B



PANEL C

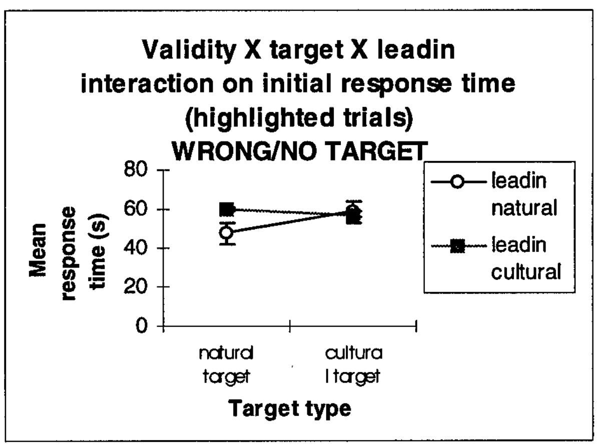


Figure 3.10: Validity X lead-in X Target interaction

<u>Confirmation time</u> A four-way analysis of variance for confirmation time, shown in Table 3.8, revealed a significant benefit for cultural targets (mean= 0.91) over natural targets (mean=1.76), possibly as a trade-off for the longer initial response times for cultural targets, as discussed in the valid trials analysis of confirmation time. When pilots initially responded more quickly to one feature type, the confirmation time is slower for that feature type since the pilot must fly longer to reach the target. If the

pilot initially responds later to one feature type, the confirmation time is faster since the pilot is closer to the object itself. However, it is not clear why this effect occurred for only lead-in features in the valid trials analysis and for only targets in the highlighted trials analysis.

Source	SS	DF	MS	F	Sig of F
Validity	8.06	2	4.03	0.64	0.526
Lead-in	12.97	1	12.97	2.07	0.151
Target	26.08	1	26.08	4.17	**0.042**
Highlighting	5.17	1	5.17	0.83	0.364
Validity X Lead-in	6.47	2	3.24	0.52	0.597
Validity X Target	12.89	2	6.44	1.03	0.359
Validity X Highlighting	9.19	2	4.60	0.74	0.481
Target X Highlighting	1.91	1	1.91	0.31	0.581
Lead-in X Highlighting	0.05	1	0.05	0.01	0.932
Target X Lead-in	16.07	1	16.07	2.57	0.111
Validity X Highlight X Target	23.53	2	11.77	1.88	0.155
Validity X Highlight X Lead-in	68.67	2	34.33	5.49	**0.005**
Validity X Target X Lead-in	3.67	1	3.67	0.59	0.444
Highlight X Target X Lead-in	5.38	1	5.38	0.86	0.355
Validity X Highlight X Target X Lead-in	32.93	1	32.93	5.27	**0.023**
Error	1206.60	193	6.25		

Table 3.8: Analysis of variance (lag time)

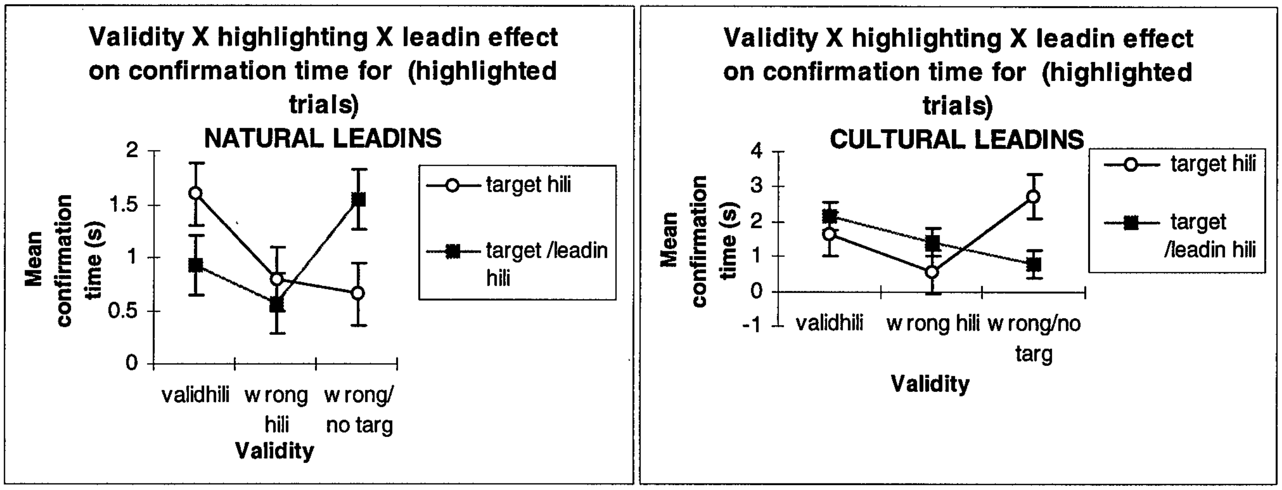


Figure 3.11: Validity X Highlighting X Lead-in interaction

The ANOVA on confirmation time also revealed a highly significant three way interaction of validity by highlighting by lead-in feature, which is shown in Figure 3.11. This interaction is the opposite of that shown in Figure 3.8. As was shown in Figure 3.8, the highlighting of the lead-in feature appears to improve accuracy when the target is missing, if the lead-in is a cultural feature but not if it is a natural features. This selective advantage in score for highlighting the cultural lead-in appears to be coupled with advantages of decreased confirmation time as well (Figure 3.11). When comparing Figure 3.11 to Figure 3.8, it appears that pilots responded more slowly under the same conditions in which they were less accurate. The cause of this relationship between score and confirmation time is unclear.

Post-experiment questionnaire

Figure 3.12 shows the pilots' subjective ratings of the different highlighting techniques. Subjects were asked to report which highlighting techniques were helpful and which were distracting. Despite the lack of significant highlighting benefits when highlighting is valid and the costs of invalid highlighting, most subjects felt that the highlighting was helpful. When asked to rate which highlighting technique was most helpful, 33% said target highlighting was helpful, and another 33% replied that both target and lead-in features were helpful. Also, more than 66% of the subjects did not find either technique distracting.

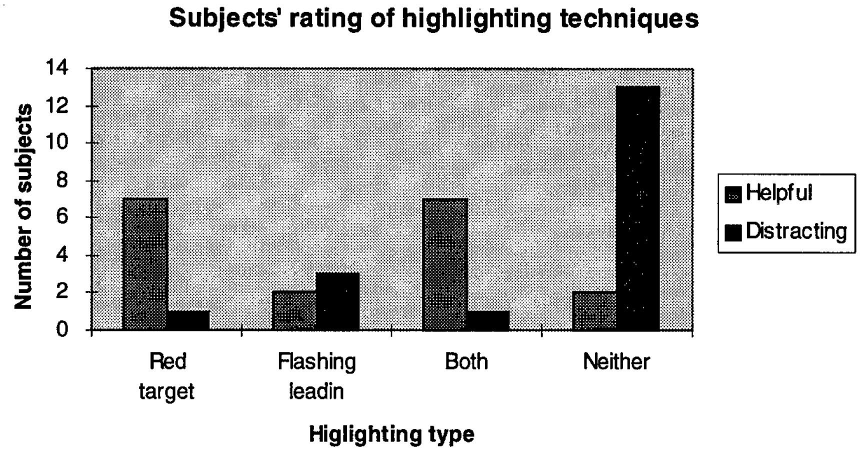


Figure 3.12: Subjective ratings of highlighting techniques

Subjects indicated that the scenes depicted in the Evans and Sutherland world became familiar no later than halfway through the experiment and the familiarity did make target identification and location easier. Also, subjects indicated that the three most useful lead-in features were lakes, rivers, and roads/highways (Figure 3.13).

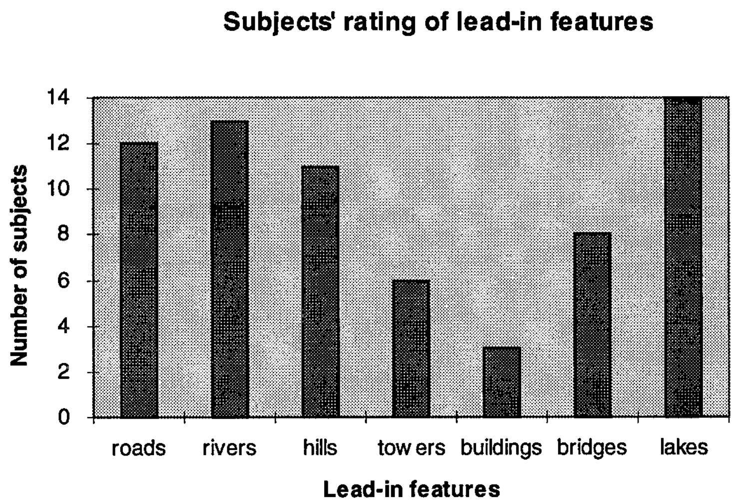


Figure 3.13: Subjective preference of lead-in features used in this experiment

Discussion

The present experiment had three primary goals in examining the dynamic target acquisition process: first, we wished to replicate the conclusion regarding the benefit of cultural over natural feature types reported by Hickox and Wickens (1996) for navigational checking tasks, and extend this to a dynamic flight situation. Secondly, since targets may be either natural or cultural , we wished to determine if valid highlighting would have any effect on the target acquisition task in a complex visual field, either by compensating for the limitations of natural features if they were found less beneficial than cultural features, or improving the performance for cultural features even above performance when natural features are highlighted. In addition, we wished to determine if highlighting targets significantly improved the search task across both feature types when compared to performance without target highlighting. Previous literature suggested that valid highlighting should significantly improve the target acquisition task performance by automating the feature search task on the map, thereby reducing head-down time spent on scanning the map, and cueing attention to a particular area of the FFOV (Fisher et al., 1989; Fisher and Tan, 1989; Donner et al., 1991). Our third goal was to examine the effects of invalid highlighting on the target acquisition task. The intelligence (human or computer automation) upon which target highlighting is based can not be assumed to be 100% accurate. Therefore, we were interested in determining if the invalid highlighting would produce "garden path" complacency errors such that pilots would not confirm that the highlighted target was correct, relying instead on automation-induced confirmation biases and top-down processing (Wickens, 1992; Singh et al., 1993). If this complacency did occur, we asked whether it could be counteracted by making a nearby lead-in feature salient, through always valid highlighting, to serve as a cue for the correct target's location. This highlighted lead-in feature would be an attempt to provide some bottom-up information to offset the possible top-down confirmation biases.

Cultural vs. Natural Feature Types

The results of this experiment suggest a small advantage for cultural targets and lead-in features over natural targets and lead-in features. Examples of this benefit can be seen in the significant main effect of target feature type in Table 3.2 and 3.8 and the main effects of lead-in feature type in Tables

3.3, 3.6, 3.7 and interactions with other variables in Figures 3.1, and 3.9. These benefits for cultural features were not large, averaging a 6% increase in accuracy for cultural targets over natural targets and 8% increase in accuracy for cultural lead-in features over natural lead-in features. Pilots demonstrated a faster initial response time of approximately 4.4 seconds for cultural lead-in features and were approximately 0.5 seconds faster in confirmation time for cultural targets. This conclusion replicates that found by Hickox and Wickens (1996) in a same-different navigational checking task. The implication here is that cultural features by their nature are more symmetrical, familiar, and therefore, more viewpoint invariant, resulting in faster identification (Biederman and Gerhardstein, 1993; Pizlo and Salach-Golyska, 1995). This advantage alone was enough to create the differences in performance between the two feature types. Highlighting the target did not appear to offer any benefit for cultural targets, and on target highlighted trials the use of natural lead-in features appeared to suffer (Tables 3.3, 3.4, and Figure 3.1).

The most significant effects of feature type were observed in the target by lead-in feature interactions which were significant for both the valid trials analysis and the highlighted trials analysis (Figures 3.2 and 3.7) . Carmody-Bubb and Dunn (1996) reported the importance of "back-up cues," such as nearby lead-in features, to target acquisition, since pilots who reported using these cues were most likely to find a correct target. Also, Snyder (1973) noted that the target location cues are as important to the target acquisition task as the target itself. These two studies suggest that the lead-in features play a significant role in the target acquisition process when they provide information about the location of the target with respect to its surroundings. In our study, this location information appears to be facilitated by a target by lead-in condition where the lead-in feature was the opposite feature type of the target. Pilots responded faster and more accurately to conditions where both feature types occurred, one as the target and the other as the lead-in. This interaction was most unexpected, but does suggest useful information about how pilots extract information from the FFOV to determine their location in space and match the auxiliary map display to the view outside the cockpit for navigational checking and target acquisition tasks.

Battiste and Delzell (1991) report that pilots use continuous (linear) features for general location information and as boundaries for an area. Discrete, specific features are used to determine exact location. Although natural features can be discrete features, such as hills, lakes, ridges, etc., even these features have a more continuous area and a tendency toward defining spatial areas. Point features, such

as hills, also have linear properties, such as bends, corners, and intersections, which may also define spatial areas (Schulte and Onken, 1995). For example, a hill can be a boundary for a terrain area, such as the area between two hills or between a river and the hill, or it can be an area in itself. The hill contains an area within its own shape and boundaries that, due to its size, can be occupied by other discrete points such as antennas, buildings, etc. which may be identifiable as separate from the hill. Therefore, pilots may use any natural feature to define an area and cultural features to determine exact location. Although pilots are most accurate when the lead-in feature type is natural and the target is cultural (Figure 3.7), either feature type can occur as the target or lead-in. The key aspect of this interaction is to have both feature types represented. The benefits of having both feature types available to the task are seen whether the target is natural and the lead-in cultural, or vice versa (see Figures 3.2 and 3.7).

The use of cultural features to determine a location suggests that pilots may use a comparison strategy proposed by Goldberg et al. (1992) to describe how topographic surfaces and maps are compared. Goldberg and colleagues suggest that people may encode features that are equally recognizable at all orientations (i.e., cultural features) first, then use these features to make the same-different determinations. In this case, cultural features would be invariant anchors to compare the scene presented on the map with that of the FFOV. The initial encoding of cultural cues could explain the slight benefit for cultural features in the navigational checking and target acquisition tasks, as well as explaining the greater cost for natural target features with natural lead-in features shown in our study's target by lead-in interactions (Figures 3.2 and 3.7). However, cultural features work best when a natural feature is also used, suggesting that while viewpoint invariant cultural-cues are the mainstay of the comparison process, they are most useful when an environment-defining natural cue is also used to maintain a more global geographical awareness. The need for a natural feature is reflected in the cost in performance for cultural target features with cultural lead-in features as compared to cultural target features with natural lead-in features. Battiste and Delzell (1991) also note the importance of natural cues for maintaining such geographical awareness when the environment is unfamiliar.

Highlighting vs. No Highlighting

Contrary to previous expectations, valid target highlighting did not significantly improve performance on the overall target acquisition task. Although pilots did report finding the target highlighting helpful (Figure 3.12), the only improvement over non-highlighted maps for this task was the

subjective confidence pilots reported in their correct target identifications (Table 3.2). According to the valid trials analyses of variance for all three dependent variables, neither accuracy nor speed increased when targets were validly highlighted compared to those not highlighted, nor did time of detection and identification show a significant benefit. We expected an improvement in search time and accuracy because according to Fisher and Tan (1989) when the visual field is complex and validity is greater than 50% subjects demonstrate a benefit for highlighted conditions. Also, since further analyses revealed subjects were not completely accurate under nonhighlighted conditions (approximately 80% accuracy), we expected that highlighting would yield some significant benefits to improve this "baseline" accuracy. However, although 60% of the total trials (including "control" nonhighlighted trials) were valid, under highlighted conditions (without the nonhighlighted "control" trials) the validity level is only 40%. Therefore, as was determined by Fisher and Tan, subjects might not have obtained a significant benefit from attending first to the highlighted option. On the other hand, since the highlighting in this experiment also included lead-in features, which were always valid, then the ratio of validly highlighted objects (both valid targets AND all lead-ins) to total highlighted objects and the extent to which subjects trusted highlighting itself may have increased the subjects' perceived validity ratio to well over 50%. This argument may explain why the pilots in our study did demonstrate a nonsignificant improvement under valid highlighting conditions, although without the benefits we expected.

While our results were similar to those found in the Fisher and Tan (1989) study, they were also similar to those reported by Wickens (1992) concerning the signal detection theory in medical domains. For example, investigations examining tumor diagnosis by radiologists determined that cueing the radiologists' attention to areas where a tumor is likely to occur did increase the likelihood of a tumor's detection, but did so by shifting response bias toward more risky behavior (more false alarms) rather than by increasing sensitivity, or discriminability from distractors. In our experiment, highlighting the target may have served as a cueing device, rather than increasing target saliency, thus shifting the pilots' response biases toward selecting the highlighted option, at the cost of some inaccurate responses (false alarms). This signal detection interpretation may explain the relatively constant confidence measure despite decreasing accuracy across the levels of highlighting invalidity (Figure 3.4). The shift in response bias could also be an indicator of or perhaps a contributor to the complacent behavior we address later.

An additional possible explanation for our results stems from the properties of the electronic map itself. In our case, the dynamically updating, three-dimensional map and its realistic color applications (blue lakes, green fields, etc.) and elevation information may have decreased the complexity of the map itself, as compared to 2D or static maps. Donner et al (1991) also found a non-significant benefit for highlighted over nonhighlighted displays when the display was more intuitively organized, such as a 3D map for our purposes, where a benefit for highlighting did exist when the display was poorly formatted, such as a 2D map. In addition, target locations may not have been sufficiently complex since the areas seldom had an extreme number of similar distractors (usually no more than three) near the correct target. These explanations are plausible for our task, since subjects were approximately 80% accurate even without highlighting, and the non-significant 8% increase in accuracy for highlighted conditions was purchased at a time cost. Although the time costs were not significant, trials with valid highlighted conditions suffered approximately 2 seconds loss for initial response time (mean=51.2 seconds) and a 0.1 second loss for confirmation time (mean=1.59 seconds) compared to nonhighlighted conditions (initial response time=49.4 seconds; confirmation time=1.46 seconds). Fisher and Tan (1989) note that when displays are simplified, the cost of ignoring the highlighting and performing a simple serial search is relatively small. Therefore, a highlighting benefit might have been found if the map was two-dimensional or static. However, since such maps are generally less useful for navigation and global awareness (Wickens, 1997), simply implementing a 3D, updating map may sufficiently simplify the visual field to eliminate the need for highlighting. The possible exception to this case may be where the target area is very complex, with numerous similar distractors. However, as discussed earlier, these conditions provide greater opportunity for premission intelligence to have selected the incorrect target (highlighting invalidity), an event which raises some important concerns that we address below.

Additionally, although some benefits were found for highlighting cultural lead-in features, the lack of benefit for natural lead-in features (Figure 3.1) implies that unless all targets and lead-ins are cultural, and the above discussion suggests this may not be an ideal situation, highlighting is largely ineffective for enhancing feature type identification performance.

Highlighting Invalidity

While the previous analysis suggests that valid highlighting produced no significant benefit for the target acquisition task, perhaps the most important message from this experiment is the cost of invalid target highlighting. Invalidity produced the most significant costs in score, due to decreased accuracy (Figure 3.4), and was a factor in almost every other interaction resulting in costs for detection and confirmation of targets. In addition, no mitigating factors, such as validly highlighting an additional lead-in feature or use of any particular feature type, could sufficiently counteract the costs of invalidity. Although many of the interactions for the highlighted trials analyses are not easily interpreted, the interactions do suggest that some measures may help counterbalance the effects of invalid highlighting. Figures 3.5 and 3.6 indicate that accuracy for cultural targets and lead-in features may not be as affected as natural targets and lead-ins by invalid highlighting. Figure 3.8 suggests that accuracy may be improved by highlighting a cultural lead-in feature rather than just the target when the incorrect target is highlighted and the correct target is absent from the FFOV. However, this improvement does not extend to the highlighting of natural lead-in features for reasons which are unclear. Also, the benefits described here are obtained at a cost in detection and identification time, as shown in Figure 3.11. Unfortunately, none of these interactions reflect a sufficient improvement for invalid trials that would minimize the effect and implications of the main effect of validity on score.

The three-way interaction of validity, target type and lead-in feature type on initial response time (Figure 3.10) depicts a unique effect of target absent trials. When the target is present in the FFOV (even if an incorrect target is highlighted), the target feature type by lead-in feature type interaction is similar to that discussed above and portrayed in Figures 3.2 and 3.7, where performance is best when the target and lead-in features are opposite feature types. However, in the target absent trials, this interaction is reversed such that subjects respond faster when the target and lead-in feature types are congruent (i.e., natural target and natural lead-in). Further analyses suggest this reversal may simply reflect a speed-accuracy trade-off that does not exist for the other two validity types. For example, pilots respond faster in target absent trials when the target and lead-in feature types are both natural; however, this condition also yields the least accurate responses.

Additionally, under target absent conditions, highlighting a cultural lead-in (always valid) both improves accuracy (Figure 3.8) and decreases confirmation time (Figure 3.11), as compared to trials where only the target is incorrectly highlighted. Unfortunately, this valid highlighting of lead-in features

does not yield the same benefits when the lead-in feature type is natural. These interactions suggest that the absence of the correct target in the FFOV (and hence, a mismatch between map and FFOV) induce a different kind of behavior from conditions where the highlighting is incorrect, but the correct target is present in the FFOV. Although scores did not significantly differ between the two types of highlighting invalidity, the differing effects of highlighting and feature type suggest the cause of inaccuracy may reflect different failures in the navigational checking process under automated conditions.

It is important to note that the results contained a number of higher level interactions whose interpretations are not readily apparent. Although we have attempted to interpret some of these interactions above, it is quite possible that some of the effects were due to the specific features picked on particular trials, rather than generic feature types. This possible influence of specific features was an inevitable by-product of the restricted navigational area used in this experiment. The restricted area limited our ability to randomly sample large numbers of features, and the effects may surface in the higher level interactions.

The validity effect on score is the most important aspect when considering the cost of invalidity because the target acquisition task generates more negative consequences from incorrect responses than from increased time of detection and confirmation. Whereas an increase in detection and confirmation time can be dangerous, especially in military bombings over hostile territory, these costs are not nearly so dangerous as bombing an incorrect target. Also, in other tasks, the costs of landing at an incorrect airport are likely greater than those of delayed identification of the correct airport.

In this experiment, subjects did fall victim to "garden path" complacency errors when the highlighting was invalid; often responding just as quickly as when the highlighting was valid. Most disturbingly, in target absent trials pilots were more likely to select an incorrect target with generally high confidence rather than indicate the target was absent with low confidence despite being informed that the highlighting may not always be correct and that in some cases, the correct target did not exist in the FFOV. Pilots were told that the correct response in target absent cases would be to call "abort" indicating the correct target was not present. The time costs for target absent trials (depicted in Figures 3.9, 3.10, and 3.11) suggest greater initial uncertainty that the incorrect target, which was highlighted on the map and was visible in the FFOV, was the correct target as compared to the target present conditions. This uncertainty stemmed from the mismatch between the map, which displayed the expected target according to the preflight map description, although not highlighted, and the FFOV, in which the correct

target did not appear. This mismatch between the map and FFOV (which did not occur when the target was present in the FFOV), should have provided some bottom-up information that the highlighting may be invalid. Unfortunately, most pilots chose to rely on the highlighting for target selection despite their own initial uncertainty, and usually reported high confidence in their incorrect target selections (as depicted in Figure 3.4). Also, since pilots were given no feedback about the accuracy of their responses, they had no reason to alter this behavior.

Past research indicates that pilots' attitudes and trust in the automation are a deciding indication of whether the pilot will exhibit complacent behavior (Wickens, 1992; Moray and Lee, 1992; Singh et al., 1993). In our experiment, we introduced a level of distrust in the automation by indicating that cases of invalidity would occur in some of the trials. Nevertheless, pilots still fell victim to complacent behavior. However, this behavior may reflect a greater distrust in personal skill rather than overtrust in the automation. In presenting Moray and Lee's (1992) model of trust, we note that trust in automation decreases as automation becomes less reliable. On the other hand, the trust in automation grows if people doubt their own skills. It is possible that in our experiment, pilots doubted the accuracy of their own memories or knowledge about the correct target appearance and location. Since pilots were given very little time (10 seconds) for preflight planning, pilots may not have formed a clear spatial array of those features which would be encountered along the flight path. In such a case, the pilots may have distrusted their own mental model more than the possibly invalid highlighting. Although the data did not reveal the decrease in confidence we had expected for these conditions (Figure 3.4), the effect of trust is still a plausible explanation for pilots' reliance on the highlighting automation. Also, since pilots were not given any feedback about the accuracies of their responses, they may not have been aware of any inaccuracies in their search process and responses, and therefore had no cause to change their target acquisition behavior. Perhaps if such feedback was provided to the pilots, their target acquisition behavior may have reflected some attempts to improve their performance, either through increased response and confirmation times, decreased subjective confidence reports, and/or more accurate target selection and abort behavior. The lack of feedback is a limitation of this study that may be remedied in other experiments to determine the extent to which the pilots' behavior was affected by their levels of trust or complacency. Still, whether overtrust in automation or distrust in personal skills, this risky behavior under uncertain conditions carries disturbing implications for automation failure under high workload conditions when the system is usually reliable.

Conclusion

In conclusion, our results reveal a negligible benefit for valid highlighting over nonhighlighted displays and an overriding cost of invalid highlighting across all dependent measures. In this experiment, the low level of validity for highlighted trials may have precluded any significant benefit for highlighting air-to-ground targets. Also, the nature of the simulated Evans and Sutherland "world" and the benefits of the three-dimensional dynamic map may have simplified the task enough to make highlighting unnecessary. Perhaps a benefit for highlighting may be found at greater validity levels and for more complex environments and tasks using less sophisticated auxiliary maps. However, from our results, we suggest improving the map itself and utilizing a cultural by natural interaction paradigm for selecting lead-in features according to target type may be effective for improving target acquisition performance. In addition, if highlighting is employed, it is imperative to ensure highlighting is valid and back-up cues are utilized. Although our experiment did not reveal a significant benefit for highlighting lead-in features, perhaps other methods of making these cues more salient, such as a different highlighting type, may help provide better bottom-up information to prevent "garden path" behavior. Also, further study of the mixed target and lead-in feature type benefits should be conducted to determine if this benefit exists for other populations, trials, and tasks and is a result of the generic properties of the two feature types, rather than a effect of the particular set of cultural and natural features selected in this experiment. If this interaction is a consistently beneficial one, further studies may determine how to capitalize on the relationship. However, unless a reasonable benefit for highlighting is found, as well as a helpful countermeasure for invalidity, we recommend using highlighting techniques to automate the air-to-ground target acquisition process only if the accuracy of intelligence can be guaranteed. We hope the conclusions reached in this experiment about these issues of feature type, highlighting, and invalidity provide insight to the navigational checking and target acquisition tasks and provide a useful framework for the future design and implementation of electronic maps.

References

Antunano, M.J., Mohler, S.R., and Gosbee, J.W. (1989). Geographic disorientation: Approaching and landing at the wrong airport. *Aviation, Space, & Environmental Medicine,* **55**, 996-1104.

Aretz, A.J. (1991). The design of electronic map displays. *Human Factors,* **33**(1), 85-101.

Battiste, V., and Delzell, S. (1991). Visual cues to geographical orientation during low-level flight. *Proceedings of the 6th International Symposium on Aviation Psychology.* Columbus, OH: Department of Aviation, Ohio State University.

Biederman, I. and Gerhardstein, P.G. (1993). Recognizing depth-rotated objects: Evidence and conditions for three-dimensional viewpoint invariance. *Journal of Experimental Psychology: Human Perception and Performance,* **19**(6), 1162-1182.

Brown, T.J. (1991). Visual display highlighting and information extraction. *Proceedings of the Human Factors Society 35th Annual Meeting.* Santa Monica, CA: Human Factors Society.

Carmody-Bubb, M.A., and Dunn, R.S. (1996). *Situational awareness and mission rehearsal training systems: An implicit approach.* Patuxent River, ML: Naval Air Warfare Center.

Carter, R.C., and Cahill, M.C. (1979). Regression models of search time for color coded information displays. *Human Factors,* **21**(3), 293-302.

Christner, C.A. and Ray, H.W. (1961). An evaluation of the effect of selected combinations of target and background coding on map-reading performance -- experiment V. *Human Factors,* **3**, 131-146.

Converse, S.A., Kozai, S., and Batten, D. (1992). Color coding to facilitate performance of focused attention tasks with object displays. *Proceedings of the Human Factors Society 36th Annual Meeting.* Santa Monica, CA: Human Factors Society.

Crampton, J. (1992). A cognitive analysis of wayfinding expertise. *Cartographica,* **29**(3), 46-65.

Donner, K.A., McKay, T., O'Brien, K.M., and Rudisill, M. (1991). Display format and highlighting validity effects on search performance using complex visual displays. *Proceedings of the Human Factors Society 35th Annual Meeting.* Santa Monica, CA: Human Factors Society.

Drury C.G., and Clement, M.R. (1978). The effect of area, density, and number of background characters on visual search. *Human Factors,* **20**, 597-602.

Eriksen, C.W., and Webb, J.M. (1989). Shifting of attentional focus within and about a visual display. *Perception and Psychophysics,* **45**(2), 175-183.

Eley, M.G. (1988). Determining the shapes of land surfaces from topographic maps. *Ergonomics,* **31**, 355-376.

Fisher, D.L., Coury, B.G., Tengs, T.O., and Duffy, S.A. (1989). Minimizing the time to search visual displays: The role of highlighting. *Human Factors,* **31**(2), 167-182.

Fisher, D.L. and Tan, K.C. (1989). Visual displays: The highlighting paradox. *Human Factors,* **31**, 17-30.

Goldberg, J.H., MacEachren, A.M., and Korval, X.P. (1992). Mental image transformations in terrain map comparison. *Cartographica,* **29**(2), 46-59.

Haber, R.N. (1987). Why low-flying fighter planes crash: Perceptual and attentional factors in collisions with the ground. *Human Factors,* **29**(5), 519-532.

Hackworth, D.H. (1991, November 18). Killed by their comrades (friendly fire). *Newsweek,* **118**, 45-46.

Hickox, J.C., and Wickens, C.D., (1996). *Navigational checking: A model of elevation angle effects, image complexity, and feature type.* University of Illinois Institute of Aviation Technical Report (ARL-96-4/NAWC-ONR-96-1). Savoy, IL: Aviation Research Laboratory.

Kaptein, N.A., Van der Heijden, A. H. C., and Theeuwes, J. (1995). Search for a conjunctively defined target can be selectively limited to a color-defined subset of elements. *Journal of Experimental Psychology: Human Perception and Performance,* **21**(5), 1053-1069.

Kleiss, J.A., (1995). Visual scene properties relevant for simulating low-altitude flight: A multidimensional scaling approach. *Human Factors,* **37**(4), 711-734.

Lee, J., and Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics,* **35**(10), 1243-1270.

MacDonald, W.A., and Cole, B.L. (1988). Evaluating the role of color in a flight information cockpit display. *Ergonomics,* **31**(1), 13-37.

Madden, D.J., and Allen, P.A. (1989). Amount and duration of attentional demands during visual search. *Perception and Psychophysics,* **45**(6), 577-585.

Martens, M., and Wickens, C.D. (1995). *Lowlighting solutions to display clutter.* University of Illinois Institute of Aviation Technical Report (ARL-95-11/NASA- 95-4). Savoy, IL: Aviation Research Laboratory.

Müller, H.J., Heller, D., and Ziegler, J. (1995). Visual search for singleton feature targets within and across feature dimensions. *Perception and Psychophysics,* **57**(1), 1-17.

Nagy, A.L., and Sanchez, R.R. (1992). Chromaticity and luminance as coding dimensions in visual search. *Human Factors,* **34**(5), 601-614.

Pizlo, Z. and Salach-Golyska, M. (1995). 3-D shape perception. *Perception and Psychophysics,* **57**(5), 692-714.

Scanlan L.A. (1977). Target acquisition in realistic terrain. *Proceedings of the Human Factors Society 21st Annual Meeting.* Santa Monica, CA: Human Factors Society.

Shontz, W.D., Trumm, G.A., and Williams, L.G. (1971). Color coding for information location. *Human Factors,* **13**(3), 237-246.

Schreiber, B., Wickens, C.D., and Alton, J. (1995). *Navigational checking: The influence of 3D map rotation and scale.* University of Illinois Institute of Aviation Technical Report (ARL-95-9/NAWC-95-1). Savoy, IL: Aviation Research Laboratory.

Schulte, A., and Onken, R. (1995). *Modeling of pilot's visual behavior for low-level flight.* Synthetic Vision for Vehicle Guidance and Control. Orlando, FL: AeroSense.

Schultz, E.E. Jr. (1986). *Decluttering methods for computer-generated graphic displays.* California Institute of Technology Invention Report (NPO-16733/6247). Pasadena, CA: Jet Propulsion Laboratory.

Singh, I.L, Molloy, R., and Parasuraman, R. (1993). Automation-induced "complacency": Development of the complacency-potential rating scale. *The International Journal of Aviation Psychology,* **3**(2), 111-122.

Snyder, H.L. (1973). *Dynamic visual search patterns.* Committee on Vision: Division of Behavioral Sciences National Research Council. Washington, D.C.: National Academy of Sciences.

Spiker, A., Rogers, S.P., and Cincinelli, J. (1986). Selecting color codes for a computer- generated topographic map based on perception experiments and functional requirements. *Ergonomics,* **29**(11), 1313-1328.

Stiff, J. (1993). Natural scene templates as cueing devices. *Proceedings of the 7th International Symposium on Aviation Psychology.* Columbus, OH: Department of Aviation, Ohio State University.

Tan, K.C., and Fisher, D.L. (1987). Highlighting and search strategy considerations in computer-generated displays. *Proceedings of the Human Factors Society 31st Annual Meeting.* Santa Monica, CA: Human Factors Society.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett memorial lectures. *Quarterly Journal of Experimental Psychology,* **40A**(2), 201-237.

Treisman, A., and Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance,* **16**, 459-478.

Van-Orden, K.F., Divita, J., and Shim, M.J. (1993). Redundant use of luminance and flashing with shape and color as highlighting codes in symbolic displays. *Human Factors,* **32**(1), 61-77.

Wang, Q., Cavanagh, P., and Green, M. (1994). Familiarity and pop-out in visual search. *Perception and Psychophysics,* **56**(5), 495-500.

Whitaker, L.A., and Cuqlock-Knopp, G. (1991). Use of visual cues by orienteers: An analysis of interview data. *Proceedings of the Human Factors Society 35th Annual Meeting.* Santa Monica, CA: Human Factors Society.

Wickens, C.D. (1992). *Engineering psychology and human performance* (2nd ed.). New York: Harper Collins.

Wickens, C.D. (1997). Frame of reference for navigation. In D. Gopher and A. Koriat (Eds.), *Attention and performance, Vol. 16.* Orlando, FL: Academic Press.

Wickens, C.D., Schreiber, B., and Renner, G. (1994). *3-D navigational checking: A conceptual model and an experiment.* University of Illinois Institute of Aviation Technical Report (ARL-94-9/NAWC-94-1). Savoy, IL: Aviation Research Laboratory.

Williams, H.P., Hutchinson, S., and Wickens, C.D. (1996). A comparison of methods for promoting geographic knowledge in simulated aircraft navigation. *Human Factors,* **33**, 355-376.

Wolfe, J.M. (1994). Visual search in continuous, naturalistic stimuli. *Vision Research,* **34**(9), 1187-1195.

Yeh, M. and Wickens, C.D. (1997). *Effects of color coding, intensity coding, and decluttering on visual search in electronic map displays.* University of Illinois Technical Report. Urbana, IL: Beckman Institute for Advanced Science and Technology.